

Articles in the world's languages

Laura Becker

University of Freiburg

April 23, 2021

Contents

1	Introduction	1
1.1	Previous work	2
1.2	The present study	8
1.2.1	Objectives	8
1.2.2	In a nutshell	10
1.2.3	Data and methodology	18
2	Articles as a crosslinguistic category	28
2.1	The problem	28
2.2	Defining articles	32
2.2.1	Articles as a crosslinguistic category	32
2.2.2	Articles encode referential functions	33
2.2.3	Articles occur in the nominal domain	37
2.2.4	Articles are systematically distributed	42
2.3	Summary	50
3	Referentiality and types of articles	52
3.1	Referentiality	52
3.1.1	The discourse situation	53
3.1.2	Identifiability, definiteness, and indefiniteness	56
3.2	The referential functions	65
3.2.1	Deictic	65
3.2.2	Anaphoric	66
3.2.3	Recognitional	67
3.2.4	Establishing	70
3.2.5	Situationally unique	71
3.2.6	Contextually unique	73
3.2.7	Bridging	79

3.2.8	Absolutely unique	83
3.2.9	Specific	83
3.2.10	Nonspecific	85
3.3	Defining the article types	85
3.3.1	Articles in the definite domain	86
3.3.2	Articles in the indefinite domain	87
3.3.3	Domain-crossing articles	88
3.4	Referent types, article types, and referential hierarchies	89
3.4.1	Referent types and article types distinguished in the present study	90
3.4.2	Referential scales	92
3.4.3	Pragmatic (strong) and semantic (weak) definiteness	95
3.5	Summary	98
4	The distinction between articles and related markers	99
4.1	Demonstratives and articles in the definite domain	100
4.2	Possessives and definite articles	109
4.2.1	Udmurt	109
4.2.2	Amharic	114
4.2.3	Indonesian	115
4.3	The numeral ‘one’ and articles in the indefinite domain	117
4.4	Negative polarity items and nonspecific articles	122
4.5	Summary	126
5	Articles in the definite domain	128
5.1	Definite articles	128
5.1.1	Kaqchikel	129
5.1.2	Mokpe	134
5.1.3	Definite articles with deictic referents	140
5.1.4	Definite articles with absolutely unique referents	145
5.1.5	Definite articles with proper nouns	149
5.2	Anaphoric articles	152
5.2.1	Limbum	152
5.2.2	Komnzo	157
5.2.3	Akan	160
5.2.4	Anaphoric articles with establishing and recognitional referents	170
5.3	Weak definite articles	174
5.3.1	Fering and German	174

5.3.2	Hausa	181
5.3.3	Lakota	186
5.3.4	Urama	190
5.3.5	Ma'di	193
5.4	Recognitional articles	195
5.4.1	Oksapmin	196
5.4.2	Lavukaleve	199
5.4.3	Bininj Kun-Wok	202
5.4.4	Yankunyjtjara	204
5.4.5	Gooniyandi	207
5.4.6	Recognitional articles: summary and outlook	209
5.5	Summary	211
6	Articles in the indefinite domain	214
6.1	Indefinite articles	215
6.1.1	Tz'utujil	216
6.1.2	Sri Lanka Malay	218
6.1.3	Bonan	221
6.2	Presentational articles	224
6.2.1	Lango	225
6.2.2	Bilua	227
6.2.3	Chatino	229
6.2.4	Kashibo-Kakataibo	231
6.2.5	Presentational articles and the development of NUM > ART:INDEF	235
6.3	Exclusive-specific articles	237
6.3.1	Biak	237
6.3.2	Akan	242
6.3.3	Palula	243
6.4	Nonspecific articles	244
6.4.1	Ayoreo	245
6.4.2	Lakota	248
6.4.3	Tongan	251
6.4.4	On the source of nonspecific articles	257
6.5	Summary	263

7	Domain-crossing articles	264
7.1	Inclusive-specific articles	264
7.1.1	Bemba	265
7.1.2	Ayoreo	269
7.1.3	Tongan: Evidence for a weak inclusive-specific article?	272
7.2	Referential articles	278
7.2.1	Rapa Nui	278
7.2.2	Baure	283
7.2.3	Halkomelem	287
7.3	Summary	293
8	Articles: Crosslinguistic trends and variation	295
8.1	Overall distribution of the 10 article types	296
8.2	Areal trends	302
8.3	The absence of generic articles	308
8.4	Referential scales: Evidence from articles	313
8.4.1	A refined referential scale	313
8.4.2	The distribution of articles along the referential scale	316
8.5	A typology of articles	317
8.5.1	Results of the present study	317
8.5.2	A comparison with Dryer's typology	320
8.6	Summary	324
9	Article systems	327
9.1	Dependencies between articles	327
9.1.1	Weak definite and anaphoric articles	328
9.1.2	Nonspecific and specific articles	332
9.2	Functional overlap of articles in the definite domain	334
9.3	Functional overlap of articles in the indefinite domain	336
9.3.1	Tepehua	336
9.3.2	Basque	341
9.3.3	Maori	347
9.4	A typology of article systems	360
9.4.1	Article systems marking the definite domain	361
9.4.2	Article systems marking the indefinite domain	362
9.4.3	Article systems marking the definite and indefinite domain	364
9.4.4	Article systems with functional overlap	368

9.4.5	The crosslinguistic distribution of article systems	369
9.4.6	Evidence for a definite and an indefinite domain	372
9.5	Summary	374
10	Conclusions	376
10.1	Main findings	376
10.1.1	Article types	376
10.1.2	Articles and discourse prominence	379
10.1.3	Article systems	381
10.2	Future research	383

List of Figures

1.1	Areal coverage of the sample	19
1.2	Areal distribution of definite and indefinite articles (Dryer 2013b)	20
3.1	Discourse situation and discourse universe	53
3.2	Scale of uniqueness (Löbner 2011: 320)	97
4.1	Grammaticalization path from demonstrative to general noun marker	100
4.2	Grammaticalization path from the numeral ‘one’ to a nominal marker	118
6.1	Grammaticalization path from the numeral ‘one’ to a nominal marker	236
8.1	Distribution of the 10 article types in the sample	296
8.2	Distribution of the 10 article types across macro areas	302
9.1	Distribution of article systems in the sample	370
9.2	Distribution of article systems across macro areas	371

List of Tables

1.1	The relevant referential functions	11
1.2	Mapping the article types to their referential functions	12
1.3	Primary data collection	21
1.4	Languages from the macro area “Africa”	22
1.5	Languages from the macro area “Eurasia”	23
1.6	Languages from the macro area “Papunesia”	24
1.7	Languages from the macro area “Australia”	25
1.8	Languages from the macro area “North America”	26
1.9	Languages from the macro area “South America”	27
3.1	Traditional view of specificity and nonspecificity	84
3.2	Overview of the referential functions	90
3.3	Mapping the article types to their referential functions	91
4.1	The functional domain of demonstratives and definite articles (Himmelman 1997: 191)	101
4.2	The cut-off point between demonstratives and articles	103
4.3	Defining criteria for the articles in the definite domain	105
4.4	Third person possessive markers in (Beserman) Udmurt	111
4.5	Characteristic properties of indefinite articles (Heine 1997: 68)	119
4.6	The cut-off point between the numeral ‘one’ and articles	120
4.7	Relevant contexts for adnominal NPIs	123
5.1	The distribution of <i>ri</i> in Kaqchikel	133
5.2	Paradigm of the definite article in Mokpe	135
5.3	The distribution of <i>è/H</i> in Mokpe	139
5.4	Kaqchikel demonstratives	144
5.5	Limbum demonstratives	153
5.6	The distribution of <i>fō</i> in Limbum	157

5.7	The distribution of <i>ane</i> in Komnzo	160
5.8	The distribution of <i>no</i> in Akan	170
5.9	Definite articles in Fering	174
5.10	The distribution of the weak definite article in Fering	180
5.11	The distribution of the weak definite article in German	181
5.12	The distribution of definite, anaphoric, and recognitional articles	212
6.1	Exponents of the Biak exclusive-specific article	238
6.2	Ayoreo article exponents	245
6.3	Lakota exclusive-specific and nonspecific article paradigm	249
6.4	Tongan article paradigm	251
7.1	Bemba noun class markers	266
7.2	Ayoreo article exponents	269
7.3	Halkomelem article exponents	287
8.1	Distribution of the 10 article types in the sample	296
8.2	Distribution of definite articles across macro areas	303
8.3	Distribution of indefinite articles across macro areas	303
8.4	Distribution of anaphoric articles across macro areas	304
8.5	Third person pronouns in Australian languages with anaphoric articles	305
8.6	The distribution of (referential) markers across the referential scale	314
8.7	Coverage of definite, specific, and nonspecific referent types	316
8.8	A typology of articles	318
8.9	Dryer's preliminary typology of articles	321
9.1	Maori determiners	349
9.2	Article systems only marking the definite domain	361
9.3	Article systems only marking the indefinite domain	362
9.4	Article systems marking the definite and indefinite domain I	364
9.5	Article systems marking the definite and indefinite domain II	367
9.6	Article systems with functional overlap	368
9.7	DEF-INDEF and INSPEC-NSPEC splits in article systems	373
10.1	The 10 article types with their referential functions	377

Acknowledgments

This book is a revised version of my dissertation that I submitted in Leipzig in 2018. First and foremost, I want to thank my two supervisors, Martin Haspelmath and Barbara Stiebels. Thank you for having been such a great team of supervisors, who gave me a lot of freedom in developing my dissertation project, but who were always there to stop me from pursuing silly or unrealistic ideas, and to discuss my questions and give me valuable feedback from both a theoretical and a typological perspective. At times I felt that it was not easy to come up with an approach or an analysis that you would both support, but in the end, this combination was exactly what I needed to force me to properly motivate and argue for my decisions. I am also thankful to Martin Haspelmath for organizing the Diversity Linguistics Seminar, letting me be part of the typology community in Leipzig. I also wish to thank Klaus von Heusinger - this book would not have the form it has without his valuable and detailed comments from yet another perspective.

My thanks also go to Gereon Müller and the other members of the IGRA graduate school. I do not think I could have written this dissertation in a different place. I am grateful for the inspiring environment that you have created for us as graduate students, allowing us to work on so many different topics, encouraging us to go to conferences, summer schools, or to do fieldwork, without which I would not be the linguist that I am today. Also on a personal level, the possibility to travel literally around the world and to meet people from so many contexts different from the ones that I am familiar with was a very enriching and, even more so, humbling experience. I am extremely grateful especially for the financial support that allowed me to do fieldwork in Cameroon, Colombia and Guatemala for my dissertation project. Related to this I also wish to thank Maria Polinsky, Tess Wood and Pedro Mateo Pedro for organizing my stay in Guatemala as part of the Guatemala Field Station. Thank you, Pedro, for organizing meetings with speakers of various Mayan languages, for allowing me to visit different interesting places and to get to know so many fascinating people and stories. Examples from Kaqchikel, Tz'utujil and Q'anjobal have made it into this book, which would be very different without my trip to Guatemala in 2017.

I am also indebted to my friend and colleague Jude Nformi, without whom this book would not have its present form either. Thank you for letting me join you on your trip back to Cameroon and for making it possible for me to work on Mokpe. Thank you also for your time and patience to discuss the Limbum examples. Without your help, this book would be short of quite some examples from Bantu. I also thank my colleague and office mate Sampson Korsah for taking the time to talk me through many Akan examples while being extremely busy with his own dissertation. Many thanks go out to the various consultants who have shared their languages and insights with me! This book would not have the shape without them.

My thanks also go to my other IGRA colleagues, who made our time in Leipzig so enjoyable. Thanks to Asia (Joanna), Zoka (Zorica), Katja, Andy, Hannes, Ludger and Sampson from the first

cohort and to Jude, Siri and Yuriy from the second cohort for all those great memories! From coffee breaks in the kitchen to the dinners on “KarLi” with Matías, Asia and Andy, and to going out for a run with Ludger. Jingting, thanks for many linguistic discussions on typology and for the fun we had in Leipzig and in all those other places we discovered at conferences and beyond!

Lastly, I thank my family and my in-laws who supported me throughout the process of writing this book. Thank you for putting up with me when I was in a non-stop work mode, thanks for making sure I got some sleep and not only coffee during the two months before the submission of my dissertation. Finally, I thank my husband, Matías, for his endless support in the venture of writing this book and in life in general.

Abbreviations

1,2,3	1st, 2nd, 3rd person	AUX	auxiliary verb
ABL	ablative	AV	actor voice
ABS	absolute	BEN	benefactive
ACC	accusative	BG	background
ACTL	actual	CAUS	causative
ACTOR	actor (voice)	CD	clausal determiner
ACT	active	CERT	certainty
ADDIT	additive	CIRC	circumstantial
ADDR	addressee marker	CJ	conjoint tense
ADJ	adjective, adjectivizer	CLEFT	cleft marker
ADMON	admonition	CL	class marker, classifier
ADNOM	adnominal	CNTG	contiguous
ADVERS	adversative	COLL	collective
ADV	adverbial	COMM	common noun
AFFECT	affective	COMPAR	comparative
AFF	affirmative	COMPL	completive
AG	agentive, agent	COMP	complementizer
ALL	allative	COM	comitative
AL	alienable	COND	conditional
AM	association marker	CONF	confirmation marker
ANA	anaphoric	CONJ	conjunction
ANT	anterior	CONSEQ	consequential
AN	animate	CONS	consecutive
AOR	aorist	CONTR	contrastive
APPL	applicative	CONT	continuous/continuative
APUD	apudessive	COORD	coordinator
ART	article	COP	copula
ASSERT	assertive	COS	change of state
ATTR	attributive marker	CUSTOM	customary verbal suffix
AUGM	augmentative	CV	converb
AUG	augmented number	DAT	dative

DEB	debitive	HAB	habitual
DECAUS	decausative	HESIT	hesitation
DECL	declarative	HESTPST	hesternal (yesterdays') past
DEF	definite	HODFUT	hodiernal (today's) future
DEM	demonstrative	HODPST	hodiernal (today's) past
DEP	dependent aspect	HON	honorific
DESID	desiderative	HORT	hortative
DETR	detransitive	HSY	hearsay
DEV	devaluation	HUM	human
DIM	diminutive	IDENT	identificational
DIR	directional	IDEO	idephone
DIST	distal	IGNOR	ignorative
DJ	disjoint tense	ILL	illative
DS	different subject (switch reference)	IMM	immediate
DUR	durative	IMPERS	impersonal
DU	dual	IMPL	implicated
DWELL	dweller	IMP	imperative
EL	elative	INAL	inalienable
EMPH	emphasis	INAN	inanimate
EPIST	epistemic possibility	INCEP	inceptive
EQ	equative	INCH	inchoative
ERG	ergative	INCOMPL	incompletive
EST	established aspect	INDEF	indefinite
EVID	evidential	IND	indicative
EV	eventive	INESS	inessive
EXCL	exclamative	INFER	inferential
EXH	exhaustive	INF	infinitive
EXIST	existential copula	INSPEC	inclusive-specific
EXSPEC	exclusive-specific	INSTR	instrumental
EX	exclusive	INTERR	interrogative mood
FACT	factitive	INTJ	interjection
FIN	phrase final marker	INTR	intransitive
FOC	focus	INT	intensive
FRAME	framing, background	IN	inclusive
FRUST	frustrative	IPFV	imperfective
FUT	future	IRR	irrealis
F	feminine	ITER	iterative
GENPOSS	general possessor	LAT	lative
GEN	genitive	LNK	linker
GOAL	goal	LOC	locative

MED	medial	POL	polarity
MID	middle	POSS	possessive
MIN	minimal number	POST	postposition
MIR	mirative	POT	potential
MOD	modal marker	PRED	predicative
MULT	multiple	PREP	preposition
M	masculine	PRESENT	presentative
NARR	narrative marker, narrative register	PRESUMP	presumptive
NEARFUT	near future	PRES	presentational
NEG	negation	PRIV	privative
NFUT	nonfuture	PROG	progressive
NF	non-feminine	PROHIB	prohibitive
NHUM	non-human	PROM	prominence marker
NMLZ	nominalizer	PROP	proper noun marker
NMOD	noun modification marker	PROSP	prospective
NM	non-masculine	PROX	proximal
NOM	nominative	PRO	pronominal
NONFIN	non-finite	PRS	present
NOST	discourse particle marking nostalgia	PRT	discourse particle
NPOSS	nonpossessed	PST1	past 1, recent past
NPST	nonpast	PST2	past2 , intermediate past
NSG	non-singular	PST3	past 3, remote past
NSPEC	nonspecific	PST:UW	unwitnessed past
NTR	neutral aspect	PST:W	witnessed past
NUM	numeral marker	PST	past
NVIS	non-visible	PTCP	participle
N	neuter	PURP	purposive
OBLIG	obligation marker	P	patient
OBL	oblique	QUOT	quotative
OBV	obviative	Q	question/interrogative marker
ORIG	origin (spatial value)	RC	relative clause marker
O	object marker	REAL	realis
PART	partitive	RECIP	reciprocal
PASS	passive	RECOG	recognitional
PC	paucal	RECPST	recent past
PERF	perfect	RED	reduplication
PFV	perfective	REFL	reflexive
PL	plural	REF	referential
PM	phrasal marker	RELAT	relational
POLIT	politeness marker	REL	relativizer

REMFUT	remote future	SUPER	SUPER localization
REMPST	remote past	SUP	superlative
REM	remote	S	subject
REPORT	reported	TA	tense-aspect marker
REP	repetitive	TC	topic change marker
RES	resultative	TEMP	temporal marker
SEQ	sequential	THEME	theme (voice)
SER	verb serializer	TOP	topic
SG	singular	TRI	trial
SIM	simultaneous	TR	transitive
SPC	nominal specifier	VBLZ	verbalizer
SRC	source	VENT	ventive
SS	same subject (switch reference)	VICIN	vicinity
STAT	stative	VIS	visible
STI	indirect stance marker	A	agent argument in transitive clauses
SUBJ	subjunctive	P	patient argument in transitive clauses
SUB	subordinate, subordinator	S	sole argument in intransitive clauses

Chapter 1

Introduction

Articles have interested linguists for a long time for a number of reasons: their contribution to the referential interpretation of nouns, their syntactic properties, their diachronic development from other linguistic elements, or their relation with other markers of e.g. definiteness, to name a few. However, even though there is a substantial body of work on various aspects of articles, we still seem to lack good language-independent criteria to properly define articles, which is important once we move away from well-studied languages. Related to that, we do not yet have a systematic overview of what are common and less common properties of articles from a typological point of view. To provide such an overview is the objective of this study.

There are three main issues concerning the classification of markers that could be articles. First, articles are sometimes classified as definite or indefinite even if their functions differ considerably from what we would usually call definite or indefinite articles. In many cases, this is simply because these categories are not defined in a precise way or because no better notions are available. Thus, we still lack coherent criteria to distinguish different types of articles. Second, descriptions or analyses of markers that could be articles are sometimes not treated as such based on the properties of articles in English or other European languages. Third, we find descriptions or analyses of markers as articles based on the fact that they occur together with nouns. For instance, many Oceanic languages have a marker which occurs with proper nouns (sometimes only in certain syntactic contexts) and which is traditionally called a proper or personal article.

What this shows is that there is still need for a precise set of functions that articles have and that can be used as a language-independent definition. The present study aims at closing this gap by defining articles as a crosslinguistic category and by analyzing the crosslinguistic trends of their properties.

In the last decades, similar advances have been made for a number of functionally or formally related categories, e.g. Aikhenvald (2008) for classifiers, Diessel (1999) for demonstratives, Haspelmath (1999) for indefinite pronouns, and Bhat (2004) for pronouns in general. Even though

articles may not have been discussed systematically in such detail from a typological perspective yet, there are of course a number of theoretical, comparative, and typological studies that deal with various properties of articles and that the present study builds on. In the following section, I will give a brief overview of important previous crosslinguistic work on articles.

1.1 Previous work

The earliest modern crosslinguistic studies that have to be mentioned are [de la Grasserie \(1896\)](#) and [Brandstetter \(1913\)](#). [De la Grasserie \(1896\)](#) offers an impressive overview of certain semantic and syntactic properties of articles from different languages of the world, including a discussion on “determination” on verbs, adjectives, and adverbs. [De la Grasserie](#) distinguishes articles according to their semantic function, but at the same time, he builds a typology of article systems, presenting systems with a separate definite and an indefinite article, with a single article expressing both definiteness and indefiniteness, and with a single definite article. While this study discusses an impressive number of languages from various areas of the world, especially given the time of publication, it is not without theoretical issues. For instance, [de la Grasserie](#) does not always clearly distinguish between articles and possessives, discussing possessive markers that are affixed to nouns in Hungarian as suffixed article ([de la Grasserie 1896: 310](#)). As [Krámský \(1972: 7-8\)](#) points out, the distinction between articles and demonstratives is not always clear either, and for many of the examples the sources are not properly indicated. Nevertheless, [de la Grasserie \(1896\)](#) remains a very impressive, early forerunner study of articles from a broadly comparative perspective.

[Brandstetter \(1913\)](#) is another early study which deals with articles beyond the languages of Europe. It offers a more in-depth description of articles in 44 Malayo-Polynesian languages spoken in Madagascar, Indonesia, Malaysia, Brunei, and the Philippines. Based on their meaning or function, [Brandstetter](#) distinguishes between seven types of articles: personal, kinship, definite, indefinite, general, linking, and partitive articles. He presents each of these article types with examples from various languages, including a section on how these articles interact with argument marking in different syntactic contexts. This is followed by a discussion of the diachronic origin of articles. [Brandstetter](#) notes that the definite and indefinite articles in Austronesian can be traced back to demonstratives and the numeral ‘one’, respectively, drawing parallels to equivalent developments in Indo-European. He approaches the diachrony of articles not only in terms of their forms, but also discusses how their functions developed, e.g. that the indefinite article had to develop from the numeral ‘one’ and not the other way around, making it another impressive early study of articles outside of Indo-European.

Besides those two very early studies, there are a number of important broad comparative works on the functions and syntactic properties of articles relevant to the present study, namely

Dryer (1989, 2013b,c), Himmelmann (1997), Krámský (1972), Leiss (2000), Moravcsik (1969), Schroeder (2006). As will become clear in the following chapters, many insights of the present study build on findings from those earlier works, which I will briefly present in turn here.

In her 1969 paper, Moravcsik approaches the topic of (definite) determination from a functional and typological perspective. The first part of the paper gives an overview of different linguistic strategies to convey definiteness in anaphoric uses, namely word order, different adnominal markers (e.g. articles, demonstratives, possessives), various types of pronouns, and modification of the nominal by adjectives or relative clauses. Moravcsik (1969) also mentions the crosslinguistic availability of the numeral ‘one’ to be used as an indefinite marker. The second part of her paper then aims to establish whether or not different types of articles, i.e. definite and indefinite ones, should be treated as subtypes of a single category of articles. Moravcsik argues that this is only warranted if we find evidence for different articles within languages to share formal properties such as their phonological shape, their position within the nominal domain, or their agreement behavior. Based on evidence from a variety of languages, she argues that there is not sufficient evidence for articles to share such formal properties within languages. She concludes that definite articles, often sharing formal and functional properties with demonstratives, should be treated as pronouns, while indefinite articles should be regarded as a numeral, given that they always involve quantification and often originate from the numeral ‘one’. Thus, Moravcsik (1969) is hesitant to posit a general category of articles under which different types of articles are subsumed. Even though this conclusion may not necessarily bring us much closer to a crosslinguistic definition of articles, many of the properties discussed in Moravcsik (1969) have laid an important foundation for the definition of articles as a crosslinguistic category in the present study (cf. Chapter 2).

Another study of substantial importance for a crosslinguistic comparison of articles is Krámský (1972), who aims at a general definition and theory of articles in relation to the semantic categories of definiteness and indefiniteness.¹ The first part of his study presents the relevant theoretical background for the discussion of articles and, more generally, linguistic ways of expressing definiteness and indefiniteness in the world’s languages. The reader interested in the historical origin of the concept of “articles” is referred to his Chapter 3 (Krámský 1972: 18-29), going back to the Greek Stoic Dionysius Thrax (4th century BC) up to various modern linguistic definitions and criteria from the 19th and the early 20th century. The second part of Krámský (1972), called “A typology of languages based on the occurrence of the category of determinedness vs. indeterminatedness”, is an impressive crosslinguistic overview of various types of article systems that we find in the languages of the world. Because the aim of this part is in principle

¹Krámský (1972) calls those categories “determinedness” and “indeterminatedness”.

very similar to the objective of the present study, I will describe Krámský's typology of article systems in more detail.

Krámský (1972) distinguishes 7 major types of languages according to how determinedness and indeterminateness are formally expressed. Those types are: (A) either or both determinedness and indeterminateness are expressed by independent words, (B) one of those categories is expressed by an independent word and the other one by a clitic, (C) both are expressed by clitics, (D) both are expressed formally on the noun itself, the category of (in)determinateness is expressed by (E) inflection, by (F) prosody, or left (G) formally unexpressed. Thus, the highest-order criterion for his typology is the form of the article, which at the same time is intertwined with its function and the availability of other articles in the language, i.e. whether the language has both definite and indefinite articles, only a definite, or only an indefinite one. It is also important to note that those 7 types are further divided into various subtypes. Group A, for instance, is further divided into systems with (i) a definite and indefinite article, (ii) only a definite article, (iii) only an indefinite article, and (iv) systems with definite, indefinite, and partitive articles. The first of these subgroups consists of 6 further subgroups which are distinguished on the basis of whether the articles are available in the singular and the plural, and whether the language has a single (in)definite article or more than one.

Distinguishing types of article systems in such a way, namely based on a combination of form and function criteria, creates a large number of theoretically possible types, not all of which are mentioned in Krámský (1972). It is important to emphasize, though, that this still is the most comprehensive overview of article systems until now, describing a number of different systems that go beyond the most commonly found three-way distinction between languages with a definite, an indefinite, or both definite and indefinite articles. To give an example, Krámský (1972) recognizes that there are a number of languages which have more than one article in the definite domain, mentioning the distinction between proper and common articles in a number of Austronesian languages, two functionally different types of definite articles in Dakota (Siouan) (Krámský 1972: 93), or different forms of the definite article for animate and inanimate referents in Ponca (Siouan). Out of these systems, the Dakota-type system will be taken up in detail in Chapter 5 under the guise of systems with anaphoric and weak definite articles. Krámský also describes how articles or the marking of definiteness can be bound to case marking, e.g. the use of the accusative marker with specific referents only in Turkish, or how the expression of definiteness can be built into the morphology of adjectives as in Serbo-Croatian. Therefore, the study of Krámský (1972) is not confined to articles in the strict sense but includes a brief overview of other morpho-syntactic devices with which the languages in the world express definiteness and indefiniteness.

A number of choices involved in the distinctions in Krámský (1972) may not be ideal, e.g. to merge formal and functional properties of articles to define types, and several choices to analyse

certain forms of articles as separate articles vs. different forms of a single article paradigm seem ad hoc and biased by how these articles are presented in traditional grammars rather than warranted by systematic criteria. Despite all of this, Krámský is a very important study, as it lays the grounds for a crosslinguistic comparison not only of single articles (e.g. the definite article in German vs. the definite article in Spanish) or article types (e.g. the definite vs. the indefinite article in Hungarian) but article systems, which is taken up in detail in Chapter 9 of the present study.

In his 1997 study, Himmelmann offers another crosslinguistic take on articles and related markers. The main focus of the study lies on exploring the diachronic development or grammaticalization of different types of so-called “D-elements” with nouns into article-noun syntagmas and similar types of fixed nominal syntagmas. Besides the diachronic development of such syntagmas, the study establishes different types of grammaticalized adnominal D-elements (roughly, articles), based on their syntactic properties, introducing the general distinction between phrasal articles and linking articles. The articles discussed in the present study primarily fall into the category of phrasal articles; as the syntactic properties of articles are not the focus of this study, such form-based distinctions will not be discussed further here.

With a fine-grained theoretical distinction of various types of contexts in the definite domain and with corpus data from spoken German, English, Tagalog, Indonesian, Nunggubuyu, and Usan, Himmelmann (1997: Chapter 3) shows the use of demonstratives, definite, and specific articles in various types of contexts. Given that the goal of the study is to explore the development of fixed nominal syntagmas, the study then discusses the semantic and pragmatic aspects of the grammaticalization of D-elements into definite and specific articles. More precisely, Himmelmann discusses in detail the development from demonstrative to definite article and the development from definite article to specific article. This chapter is highly relevant for the present study for two reasons.

Firstly, with the discussion of specific articles and their diachronic origin from definite articles, Himmelmann (1997) makes very explicit the need for a precise distinction of different article types beyond definite and indefinite ones. To elaborate on this and to explore in detail which article types can be found in the languages of the world is one of the main objectives of the present study.

Secondly, Himmelmann (1997) provides an in-depth overview of different types of definite contexts, emphasizing that definiteness is not a primitive semantic category but can be further subdivided into different referential functions. As will be shown in Chapter 3, I am building on those distinctions as well to define different types of referential functions in the definite domain. These will then be used for the definition of four distinct article types, which cover different parts of the definite domain. Therefore, breaking up the definite domain into different referential functions as proposed in Himmelmann (1997) is essential for the distinction of article types proposed in this study. Other important studies that have to be mentioned in this respect are Am-David

(2016), Schwarz (2009, 2013), who offer detailed crosslinguistic work on systems with two types of definite articles. Put simply, in such systems, one of the definite article is used in anaphoric contexts, while the other one covers other types of definite, (mostly) non-anaphoric uses. Besides the distinction between such “strong” and “weak” definite articles (cf. Chapter 5), also the observations of Schwarz (2009) regarding different types of bridging contexts are important for the semantic distinctions made in this study.

Leiss (2000) examines the diachronic processes that have led to the development of articles, focusing on the development of definite articles in a number of Germanic languages, and taking into account the interactions between the genesis of articles and the (loss of) marking of verbal aspect. Relevant for the present study is the typological outlook in Chapter 6 of Leiss (2000). For instance, Leiss finds evidence in Germanic, Romance, and Slavic pointing towards the origin of definite articles in rhematic contexts rather than in anaphoric contexts, which is a different yet similar finding to Himmelmann (1997) in that anaphoric contexts may not be the direct bridge between demonstratives and definite articles, as had been commonly assumed. While the details of the development of definite articles go beyond the purposes of the present study, Leiss’s work is important, as it emphasizes the need to distinguish different types of definite contexts.

Another fairly recent comparative study of articles is the one of Schroeder (2006). Schroeder provides a detailed overview of the articles and article systems found in northern, eastern, and southeastern Europe, including descriptions of North Frisian and Low German varieties with two definite articles, and of a number of Uralic and Turkic languages with definite markers which originated from possessives. Examining the variation found in Eurasia in terms of articles and the marking of definiteness, specificity, and indefiniteness, Schroeder (2006) offers the currently most detailed areal analysis of articles, discussing the areal distribution of certain properties, and including intermediate areas with emerging articles due to language contact.²

To this day, the work of Dryer (1989, 2013b,c) offers the broadest crosslinguistic perspective on definite and indefinite articles. However, Dryer (1989) investigates the correlation between the order of articles and nouns with VO and OV word orders. Therefore, his main focus does not lie on articles as such, even though it is based on the largest crosslinguistic sample of articles available to date. Similarly, Dryer (2013b,c) are two crosslinguistic studies of definite and indefinite articles which correspond to two chapters in the World Atlas of Language Structure (WALS, Dryer & Haspelmath 2013). Since this project aims at an overview of various grammatical phenomena on a crosslinguistic scale, these two chapters can offer but a very condensed overview of a few selected properties of articles. In addition, the distinction of article types is kept as minimal as possible, differentiating between definite and indefinite articles only. While this choice may have

²The focus on Eurasia and the areal perspective of Schroeder (2006) is due to the fact that it emerged from work in the EUROTYP project.

been justified given the overall aim of WALS, it resulted in coercing different article types from various languages into only two types. However, as becomes evident from previous work, other, additional types of articles have to be distinguished as well. This is also one of the main objectives of the present study, to stake out the referential space that articles in the world's languages cover (cf. Chapter 3), and to classify the attested articles into well-defined types (cf. Chapters 5 to 7).

Because Dryer's classification of articles is fairly simplified, the WALS overview in Dryer (2013b,c) has not remained without criticism, e.g. from Davis et al. (2014). In their broader discussion of methods in typology, Davis et al. (2014) use the particular case of determiners or articles in Skwxwú7mesh, a Salishan language, to show that the simplistic classification as a definite article is not motivated in this case, since the article can be used in a number of indefinite contexts as well. In a reply to this criticism, Dryer (2014) clarifies a number of methodological issues, especially concerning the definition of articles. Therefore, it offers a discussion (albeit condensed) of various types of articles found across languages, including a description of selected peculiar properties of articles, most of which concern a more fine-grained differentiation of anaphoric and nonanaphoric definite functions, e.g. the availability of different definite articles for recent and more distant anaphoric referents. Based on the variety uncovered, Dryer proposes a reference hierarchy (cf. Chapter 3) which he in turn uses to propose a typology of articles. The issue of attested and unattested types of articles is taken up in Chapter 9, comparing Dryer's findings to the ones of the present study.

Two other important seminal studies highly relevant to the present one are Hawkins (1978) and Lyons (1999), even though their main focus does not lie on articles but on definiteness, indefiniteness, and referentiality. Hawkins (1978) offers a comprehensive theoretical approach on definiteness and indefiniteness. He distinguishes and describes a number of different types of referential functions, relating their expression to the use of articles in English. While not being typologically oriented, the fine-grained semantic distinctions lay an important groundwork for all following studies concerned with referentiality or articles including the present one. Chapter 3 of Hawkins (1978) describes in detail the anaphoric and immediate situation uses as well as the unfamiliar uses of the English definite article. The first type of contexts includes (associative) anaphoric and bridging uses, while the second mainly corresponds to establishing uses. The distinction of referential functions in the definite domain in Chapter 3 of the present study are mostly based on those uses described by Hawkins (1978). Related to this, Hawkins also addresses the question of distinguishing between demonstratives and definite articles (cf. Section 4.1). Chapter 4 of Hawkins (1978) then offers an overview of specific and nonspecific contexts in the indefinite domain. Therefore, even though focusing on English, Hawkins is a valuable contribution for later studies which are concerned with referential functions such as the present one.

In a similar way, Lyons (1999) is very important for establishing an overview of articles and article systems in the world's languages. He approaches the topic from a less theoretical but more typological perspective, discussing examples of definite and indefinite markers from a variety of typologically diverse languages. In order to define definiteness, Lyons discusses two different approaches based on uniqueness and familiarity in more detail (cf. Section 3.1). Amongst other markers of definiteness, Lyons groups definite articles into types according to their morphological properties, i.e. whether or not they are bound or free markers. While this way of classifying articles according to their form may not be very helpful for a consistent crosslinguistic typology of articles (which also holds for the early classification in de la Grasserie (1896)), it is very useful for providing an overview of the kinds of markers that need to be considered when discussing articles. Showing the crosslinguistic variation in their formal realization helps to move away from the view on articles biased by well-studied systems such as the articles in a number of European languages. He further builds a theory of definiteness, summarizing previous theories and approaches (Lyons 1999: Chapter 7).

Chapter 9 of Lyons (1999) then discusses certain diachronic aspects of (in)definiteness. This is relevant for the present study in an indirect way. Although my main focus lies on a synchronic overview of articles and article systems, an important question is how different types of articles can be distinguished from e.g. demonstratives or the numeral 'one', which are often the diachronic source for definite or indefinite articles (cf. Chapter 4). Besides Himmelmann (1997), Leiss (2000), Lyons (1999), other important studies concerned with the development of articles across languages are De Mulder & Carlier (2011) for definite articles and Heine (1995, 1997) for indefinite articles. Notable work on the development of articles (or (in)definite markers) in specific language families is e.g. Szczepaniak & Flick (2020) for German, Fraurud (2001), Gerland (2014), Nikolaeva (2003) for Uralic, and Huehnergard & Pat-El (2012), Pat-El (2009) for Semitic. The development of (potential) definite articles in Uralic is especially interesting in that the markers originate from possessives and not from demonstratives (cf. Section 4.2).

1.2 The present study

1.2.1 Objectives

As was shown in the previous section, there is a substantial body of crosslinguistic work on articles. However, to the best of my knowledge, a detailed and systematic crosslinguistic overview of articles is still not available. Therefore, the objective of the present study is to build on this previous work and to provide such a systematic overview, exploring the referential functions that are covered by articles in the world's languages, the attested combinations of articles into article

systems, and exploring the crosslinguistic tendencies, i.e. frequent, less frequent, and unattested patterns.

To do so, I propose a language-independent definition of articles as a crosslinguistic category based on three main criteria: their domain (the noun phrase), their function (marking the referential function of the nominal expression that they occur with), and their systematic occurrence. Then I define different referential functions which are used to define different types of articles. Related to that, I discuss the differences in terms of referential functions between certain types of articles and other related markers, e.g. definite articles vs. demonstratives. Using a sample of 104 languages with 141 articles, this study offers a detailed description of the 10 major article types with examples from a variety of typologically diverse languages. Then I explore the crosslinguistic trends and variation concerning the distribution of those 10 article types, including the question of why we have no crosslinguistic evidence for generic articles. After examining the overall distribution of article types in the sample and areal trends, I propose an extended referentiality scale based on the crosslinguistic data of the 10 article types, which at the same time can be taken as a typology of articles. The last part of this study is concerned with article system, i.e. with the interactions and combinations of articles within single languages. For instance, there are certain types of articles that do not occur in systems without certain other types of articles. The other focus of this part lies on article systems that include functionally overlapping articles, which are attested in the definite as well as in the indefinite domain. Especially in the indefinite domain, such patterns of functional overlap appear to be driven by discourse-pragmatic factors; three such systems are described in detail. The final major topic of this study is a typology of article systems, i.e. an overview of attested and non-attested article systems.

In other words, the main focus of the present study is the large-scale, typological overview and comparison of articles and article systems, capturing crosslinguistic trends and variation. Therefore, the aim is not an in-depth analysis of the systems of particular languages. Neither is the objective of the present study to capture all the details and additional properties of the articles discussed. Rather, the aim is to build a language-independent conceptual base for the crosslinguistic comparison of articles and their delimitation from related markers, focusing on their referential functions and most typical properties. Even though a number of case studies are provided, they only serve to illustrate and contribute to the crosslinguistic picture and they should not be taken as an exhaustive discussion of the article (system) of the language in question. Whenever more detailed descriptions or analyses of the articles described are available in the literature, I mention the relevant references.

A typological study based on a large sample of languages and on data mostly taken from grammars and language descriptions comes with certain limitations. I cannot provide detailed quantitative analyses of the use of the articles discussed. Although this would be extremely helpful

in certain cases, it would also go far beyond the scope of the present study, and future work is needed to subject the observations made here to quantitative language-specific or comparative testing.

The present study is not couched in any particular theoretical framework but loosely assumes a usage-based approach to grammar, meaning that language is assumed to be a dynamic, constantly changing network, shaped by our human cognition and systematic, language-independent processes of changes (grammaticalization).

1.2.2 In a nutshell

Chapter 2: Articles as a crosslinguistic category

Here I define articles as markers that occur systematically with lexical referential expressions (nouns) to mark the referential function of that expression. This definition includes three types of criteria. The function criterion allows us to distinguish between the main types of articles on the basis of their referential functions: definite, anaphoric, weak definite, recognitional, exclusive-specific, indefinite, nonspecific, inclusive-specific, weak-inclusive specific, and referential articles. The form criterion ensures that only markers occurring together with nouns qualify as articles, and it excludes, e.g. definiteness-sensitive adjective markers. The distribution criterion requires that articles occur systematically to indicate the referential function of the noun that they occur with. Systematicity does not mean obligatoriness in the sense that the article has to occur in all contexts in which it is expected due to its referential functions. Rather, articles can still be treated as such even though they do not occur in all contexts in which they would be referentially required, as long as the exceptions can be captured by rules or strong tendencies. I provide a number of examples of markers that I consider to be articles based on these three criteria as well as examples of markers that I do not treat as articles.

Chapter 3: Referentiality and types of articles

In order to define the relevant referential functions and article types, a number of important notions related to discourse and referentiality are introduced in Section 3.1 first. Section 3.2 then defines the relevant referential functions needed to distinguish different types of articles. A definition and an example of the relevant referential functions is given in Table 1.1. The referential functions are then used for the definition of 10 different types of articles (Section 3.3).

Table 1.2 shows a summary of how those different article types can be mapped to their respective referential functions. Definite articles (DEF) encode all relevant referential functions in the definite domain. I argue that they have to encode at least anaphoric and contextually unique uses. Anaphoric articles (ANA) primarily encode anaphoric uses. They do however often also

mark recognitional, establishing, rel-bridging, and in certain cases even situationally unique referents as well. Their counterpart in the definite domain are weak definite articles (DEF_{WEAK}), which are only used to mark situationally unique, contextually unique, and u-bridging referents. The last article type in the definite domain is the recognitional article (RECOG), which occurs with recognitional referents and is sometimes also used to establish referents in the discourse.

Table 1.1: The relevant referential functions

$\mathbf{R}_{\text{anaph}}$	An anaphoric referent is identifiable based on its shared identity with a previously mentioned referent. <i>A: Did you like the movie last night? B: The movie was great!</i>
$\mathbf{R}_{\text{recog}}$	A recognitional referent is identifiable based on experience or knowledge previously shared by the speaker and the hearer. <i>What happened to the cat (we used to have)?</i>
$\mathbf{R}_{\text{establ}}$	An establishing referent is marked as identifiable and thus constructed as such. <i>Did you hear the news? They are going to close the museum.</i>
$\mathbf{R}_{\text{rel-bridge}}$	A rel-bridging referent is identifiable based on its relation to another previously mentioned referent, allowing for an anaphoric relation between the two referents. <i>Eva bought a book. The author is French.</i>
$\mathbf{R}_{\text{sit.unique}}$	A situationally unique referent is identifiable because it is the only salient referent of its kind in the immediate discourse situation. <i>Beware of the dog.</i>
$\mathbf{R}_{\text{cont.unique}}$	A contextually unique referent is identifiable because it is constructed as the only salient referent of its kind in a larger context. <i>What's the best way to the center?</i>
$\mathbf{R}_{\text{u-bridge}}$	A u-bridging referent is identifiable based on its relation to another previously mentioned referent, which allows to construct the u-bridging referent as contextually unique. <i>I saw an old house in the village. The roof had been severely damaged.</i>
\mathbf{R}_{spec}	A specific referent is not identifiable but linked to a particular referent of its kind. <i>I met a strange new neighbour yesterday.</i>
$\mathbf{R}_{\text{nonspec}}$	A nonspecific referent corresponds to a single but no particular referent of its kind. <i>Do you have a pen? Any pen will do.</i>

In the indefinite domain, I show that we have to distinguish three types of articles. Indefinite articles (INDEF) encode both specific and nonspecific referents. Exclusive-specific articles (EXSPEC) are used to express specific referents only. Similarly, nonspecific articles (NSPEC) can only be used in nonspecific contexts. In addition, we find three more article types that express functions from both the definite and the indefinite domain. Inclusive-specific articles (INSPEC) encode the same functions as definite articles, but they are used to encode specific referents as well. A weak

inclusive-specific article ($INSPEC_{WEAK}$), like a weak definite article, is not used in pragmatic definite contexts. Finally, referential articles (REF) correspond to the broadest type of articles, encoding all referential functions distinguished in the definite and indefinite domain.

Table 1.2: Mapping the article types to their referential functions

article	definite domain							indefinite domain	
	R_{anaph}	R_{recog}	R_{establ}	$R_{rel-bridge}$	$R_{sit.unique}$	$R_{cont.unique}$	$R_{u-bridge}$	R_{spec}	R_{nspec}
DEF	✓	✓	✓	✓	✓	✓	✓		
ANA	✓	(✓)	(✓)	(✓)	(✓)				
DEF _{WEAK}					✓	✓	✓		
RECOG		✓	(✓)						
INDEF								✓	✓
EXSPEC								✓	
NSPEC									✓
INSPEC	✓	✓	✓	✓	✓	✓	✓	✓	
INSPEC _{WEAK}					✓	✓	✓	✓	
REF	✓	✓	✓	✓	✓	✓	✓	✓	✓

Section 3.4.2 is concerned with referential scales and hierarchies reflecting the structure between different referential functions. I provide a brief overview of proposals for referential scales and hierarchies from the literature, relating them to the distinction of referential functions as proposed in the present study. An especially important distinction is the one between pragmatic and semantic definiteness, discussed in more detail in Section 3.4.3.

Chapter 4: The distinction between articles and related markers

Section 4.1 deals with the distinction of definite articles from demonstratives. As was done in previous studies, I treat markers as demonstratives if they only encode pragmatic definite functions but no semantic definite ones. Definite articles are argued to encode referents of all definite contexts except for deictic ones, the core domain of demonstratives. Anaphoric articles are generally restricted to pragmatic definite contexts, but in contrast to demonstratives, they do not occur with deictic referents.

Section 4.2 aims to shed some light on the relation between definite articles and possessive markers. Reviewing data from Udmurt, I follow previous studies in concluding that Uralic possessives may have extended referential functions but should clearly not be considered as definite articles. They can occur in pragmatic definite contexts, but their occurrence in certain referential contexts rather seems to be the result of discourse-pragmatic factors and not referentiality as such.

I then briefly present Amharic and Indonesian, two other prominent examples of the development from possessive to definite article. All the observations taken together, however, I conclude that the crosslinguistic evidence for a development of possessive markers into definite articles is very weak, and it remains to be shown whether possessive markers can develop into definite articles.

In Section 4.3 I then deal with the distinction between the numeral ‘one’ and different types of articles in the indefinite domain. I argue that the distinction can be based on a scale from the functions of individuation and presentative marking, to specific and nonspecific marking. Numerals generally have the potential to assume the function of a presentative marker, i.e. to occasionally mark a referent as non-identifiable. If this use is not systematic but only emphatic and never required, the marker is not counted as an article. If a marker is systematically used to encode specific referents, it is a specific article. Indefinite articles are used to mark both specific and nonspecific referents. If the marker is used consistently to signal the non-identifiability of the referent (specific or nonspecific), even though it is restricted to discourse-prominent referent, I argue that it should still be analysed as an article. I call such articles presentational articles, and they represent a type of indefinite articles. If a marker only encodes nonspecific but not specific referents, it is analysed as a nonspecific article.

Nonspecific articles also need to be delimited from a subgroup of NPIs (negative polarity items). I argue in Section 4.4 that even though nonspecific articles and certain types of NPIs have the same referential functions, NPIs lead to pragmatic effects that nonspecific articles do not have. This manifests itself in the optionality of NPIs; there always is a simpler construction with similar referential functions available in the language (a bare noun or a noun with an indefinite article). This is not the case for nonspecific articles, whose use in nonspecific contexts is required.

Chapter 5: Articles in the definite domain

In Chapter 5, I provide an in-depth overview of the four article types in the definite domain using data from various languages of the world. For definite articles, I present two examples from Kaqchikel and Mokpe in Section 5.1. Then I briefly discuss the use of definite articles with deictic and absolutely unique expressions as well as with proper nouns. Across languages, there is no evidence for or against the occurrence of definite articles in such contexts, which is why I argue that they can but do not need to be marked by definite articles.

Section 5.2 deals with anaphoric (i.e. “strong definite”) articles. I present data from Limbum, Komnzo, and Akan. The articles in Limbum and Komnzo being more typical examples, the article in Akan has received much attention in the previous literature, and a number of different analyses have been proposed. As will be shown, the anaphoric article in Akan is difficult to classify because it is functionally more extended than most anaphoric articles, as it can be used to mark a referent as situationally unique as well. I argue that it should be analysed as an anaphoric article nevertheless,

mainly because it cannot be used to mark contextual uniqueness. In Section 5.2.4, I show that anaphoric articles are generally used in recognitional and establishing contexts as well, supporting the hypothesis that those contexts play an important role in the development of articles from demonstratives.

In Section 5.3, I then provide a crosslinguistic overview of weak definite articles, presenting the relevant examples mentioned in the previous literature: Fering, German, Hausa, Lakota, Ma'di, and Urama (from the sample of the present study). Out of these, the articles in Fering and German (besides other Germanic varieties) appear to be the only candidates of weak definite articles. The data from Hausa, Lakota, and Urama rather shows that the availability of an additional anaphoric marker or article does not necessarily restrict a definite article to semantic definite contexts. Based on this evidence, I conclude that although weak definite articles are attested, they are extremely rare. The more common scenario leads to a system with functionally overlapping articles rather than the formation of a weak definite article.

The last part of this chapter, Section 5.4, is concerned with recognitional articles. I briefly present five examples from Oksapmin, Lavukaleve, Bininj Kun-Wok, Yankunytjatjara, and Gooniyandi. This article type has received little attention in both the descriptive and the theoretical literature. Unlike previous studies that have defined recognitional markers as being used to mark shared personal knowledge, the data presented in this study suggests that recognitional articles are often used to refer to culture-specific entities or concepts. Thus, their use is not restricted to shared personal knowledge but includes cultural knowledge shared by the entire speech community. Given that all five languages of the sample that have a recognitional article are spoken by smaller, close-knit speech communities, I argue that their occurrence fits into the wider context of certain socio-linguistic factors favoring more complex pronominal and deictic systems.

Chapter 6: Articles in the indefinite domain

Section 6.1 is an overview of indefinite articles, discussing three examples from Tz'utujil, Sri Lanka Malay, and Bonan. In addition, I argue that we need to distinguish a further type of indefinite articles restricted to occur with discourse-prominent referents, which I call presentational article. Presentational articles are discussed in detail in Section 6.2 with examples from Lango, Bilua, Chatino, and Kashibo-Kakataibo. In the light of presentational articles, I revisit the diachronic development of indefinite articles from the numeral 'one', which usually assumes that the semantic extension from specific to nonspecific goes together with a pragmatic extension. The data from presentational articles, however, shows that these two developments are independent from each other.

I then present exclusive-specific articles in Section 6.3. Exclusive-specific articles are only used in contexts of specific reference, as is shown with examples from Biak, Akan, and Palula. The last

type of articles in the indefinite domain is the nonspecific article, marking referents as nonspecific. In Section 6.4, I describe the nonspecific articles in Ayoreo, Lakota, and Tongan. Especially with regard to nonspecific articles and markers in North and Central America, I show that their source can be traced back to a verbal irrealis marker.

Chapter 7: Domain-crossing articles

This chapter is concerned with articles that encode referential functions from both the definite and the indefinite domain. Section 7.1 introduces inclusive-specific articles with examples from Bemba, Ayoreo, and Tongan. Inclusive-specific articles can mark a referent as definite or specific. The Tongan article is especially interesting, since it may be analysed as a weak inclusive-specific article by analogy to weak definite articles. In other words, the inclusive-specific article in Tongan is used for semantic definite and specific reference, but it does not seem to occur with pragmatic definite or nonspecific referents. Section 7.2 is concerned with referential articles. Referential articles encode all functions on the referential scale. I present examples of this article type from Rapa Nui, Baure, and Halkomelem. In Rapa Nui, the referential article is paradigmatically opposed to a predicate marker that occurs with nominal expressions to mark them as predicates or as non-referential. Syntactically, both the article and the predicate marker should be analysed as determiners, though. Baure, on the other hand, uses copulas and other verbal morphology with nominal expressions in those cases in which the expression is used as a predicate. In Halkomelem, in opposition to referential expressions marked by the referential article, nominal expressions that are used as predicates receive no formal marking.

Chapter 8: Articles: Crosslinguistic trends and variation

This chapter is a crosslinguistic outlook of the distribution of the 10 article types. Section 8.1 first presents their distribution in the sample of the present study. Definite articles being confirmed as the most frequent type of articles in the world's languages (33% of the articles in the sample), I show that definite articles are in fact not substantially more frequent than indefinite articles (27%). The third most frequent type is the anaphoric article (16%), while the other article types each correspond to less than 10% of the sample. Then, I point towards a few areal trends that emerge from the data in Section 8.2, namely the high proportion of definite articles in Africa and North America, the high proportion of indefinite articles in Eurasia and South America, the lack of articles in Australia besides anaphoric and recognitional ones, and the concentration of recognitional articles in Papunesia and Australia. Section 8.3 comments on the crosslinguistic absence of generic articles (Universal 1).

Against the background of the article types discussed in this study, I revisit the question of referential scales in Section 8.4 and propose a revised scale, including a more fine-grained distinc-

tion of the definite domain, and extending it to non-referential contexts on the other end of the scale. The scale is shown in (1) below.

(1) **Referential scale**

deictic > anaphoric > recognitional, establishing, rel-bridging >
situationally unique > contextually unique, u-bridging >
specific > nonspecific > non-referential

I also examine the distribution of the articles from the sample along the scale, counting the types of article expressing certain functions. The results show that pragmatic definite functions are expressed by the highest number of articles in the sample, whereas nonspecific referents are expressed by the lowest. This suggests a cline of referential functions encoded by articles along the referential scale (Universal 2).

In Section 8.5 I propose a typology of article types. This typology is then compared to an earlier proposal by Dryer (2014). An important finding concerns the development and functional extension of articles that mark more than one referential function, namely anaphoric and definite articles in the definite domain, the two domain-crossing types of inclusive-specific and referential articles, as well as exclusive-specific and indefinite articles in the indefinite domain. These generally extend their functions from the “left” to the “right”, and not in the opposite direction. As a result, all definite, inclusive-specific, and domain-crossing articles also encode pragmatic definite expressions (Universal 3). Notable exceptions to this are the weak definite and weak inclusive-specific articles. I argue that such exceptions are always the result of a later restriction by an additional marker that blocks the use of a former definite or inclusive-specific article in pragmatic definite contexts. This is paralleled by nonspecific articles in the indefinite domain, which usually originate from former indefinite articles and are restricted to nonspecific contexts in a subsequent development as well. Another related—and somewhat obvious—finding is the absence of articles that would mark definite and nonspecific referents without marking specific ones (Universal 4).

Chapter 9: Article systems

Building on the data discussed in the previous chapters, I focus on dependencies between certain types of articles in Section 9.1. This concerns weak definite, weak inclusive-specific, and potentially weak referential articles which are not attested in systems without an anaphoric article. In the indefinite domain, the same holds for nonspecific articles that are only attested in languages that have an exclusive-specific or inclusive-specific articles as well.

Section 9.2 then examines different article systems that include functionally overlapping articles. Functional overlap in the definite domain is attested in systems with a definite and an

anaphoric or a recognitional article. In the case of languages with a definite and an anaphoric article, the definite article is replaced by the latter in contexts in which the anaphoric or pragmatic definite nature of the referent needs highlighting. This is similar to the interplay between definite articles and demonstratives in other languages. In the indefinite domain, I show that we find a similar situation with discourse-prominent referents which are not identifiable. In those cases, an inclusive-specific or a referential article is not sufficient to signal that the referent is not identifiable. This is where an additional presentational article comes in, highlighting that the referent is non-identifiable but at the same time relevant in the discourse. I also show that in certain cases, the presentational article is used to highlight that the speaker lacks sufficient information to unambiguously identify the referent, or to highlight that the exact identity of the referent is not relevant. I use examples from Tepehua, Basque, and Maori to illustrate this language-internal interplay of different article types in the indefinite domain.

In Section 9.4, I propose a typology of article systems. I show that we can distinguish between four groups of article systems: those that only encode referential functions from the definite domain, from the indefinite domain, from both domains, and those that include functionally overlapping articles. In total, I show that we can distinguish (at least) 22 types of attested article systems. In addition to those, I also discuss a number of theoretically interesting systems that are not attested. For the most part, I argue that we should not categorically exclude their existence, i.e. the probability of their existence is very small but not 0. Such systems involve a number of independent diachronic processes that are all very rare. Therefore, languages that combine several of these rare developments can be expected to be even rarer, and may simply escape our awareness. Systems that can be excluded are those that involve the dependencies of weak articles on anaphoric articles as well as the dependency of nonspecific articles on exclusive/inclusive-specific articles (Universals 6 and 6).

I also briefly discuss the crosslinguistic distribution of article systems. The most common systems consist of a single definite, indefinite, or anaphoric article. These three systems make up almost 60% of the sample. Systems that consist of more than one article are generally rare; the most “common” ones are systems with a definite and an indefinite or an exclusive-specific article. This observation relates to the finding about how article systems generally split the referential space. Including the more complex article systems as well, I show a crosslinguistic preference for article systems that make a distinction between the definite and the indefinite domain. An alternative attested split is to group definite and specific functions together, as opposed to nonspecific referents; such systems are less common (Universal 7).

1.2.3 Data and methodology

The major part of this study is based on a sample of 104 languages with 141 articles in total. Since I only include languages that have articles, the sample is a convenience sample to a certain extent. The sample results from a pre-sample of 150 languages, some of which were excluded at a later stage mostly because the examples in the descriptions did not offer a clear enough picture on the exact nature of the marker. Nevertheless, aiming at a faithful representation of the crosslinguistic variation, the sample contains approximately between 10 and 20 languages for each of the six macro areas used in the World Atlas of Language Structure and Glottolog (Dryer & Haspelmath 2013, Hammarström et al. 2020).³ The macro areas are: Africa, Australia, Eurasia, North America, Papunesia, and South America. The languages in my sample are provided in Tables 1.4 to 1.9 at the end of this section, sorted by macro area, language family, and listed together with the source(s) used in this study.

Figure 1.1 shows the areal distribution of the languages in my sample.⁴ The major area that the sample does not cover are most parts of Russia. This is due to the general lack of languages with articles in this area. The areal distributions of definite and indefinite articles observed in Chapter 37 of WALS (Dryer 2013a) show a similar pattern. Figure 1.2 below features the map that belongs to this chapter.⁵ Blue and red dots indicate definite articles, white diamonds correspond to languages with an indefinite article but no definite articles, and white dots stand for languages without articles. Without going into detail with respect to the different types of articles marked in the map in Figure 1.2, we see that even in a larger sample, Russia or rather north-east Eurasia in general seems to lack languages with articles.

It also follows from this unequal distribution of articles in the world's languages that the macro areas do not contain an equal number of languages; in the sample used in this study, Australia has by far the lowest number of languages in the sample, which follows from the a general trend against articles in that macro area. Since those differences most probably reflect the underlying distribution of articles in the world, it is not problematic for the purposes of the present study.

For large language families, the sample includes a higher number of languages.⁶ This is the case for Atlantic-Congo, Indo-European, Austronesian (especially Oceanic). Since the different languages from these families are generally diverse, especially with respect to their article-related properties, this is not problematic either for the purposes of the present study. My sample also includes a number of languages from the Arawakan family. The Arawakan family is not as large

³This division into six macro areas is not without its downsides, as discussed in Hammarström & Donohue (2014). Because my sample is a convenience sample constrained by the presence of articles in the world's languages, and because areal effects are not the main interest of this study, this issue can be neglected.

⁴The information about the geographic location of the languages is taken from Glottolog (Hammarström et al. 2020) and can be found in the online supplementary material of this study.

⁵The map is available online at <http://wals.info/feature/37A>.

⁶Large in the sense of a large number of languages (cf. <http://glottolog.org/glottolog/family>).

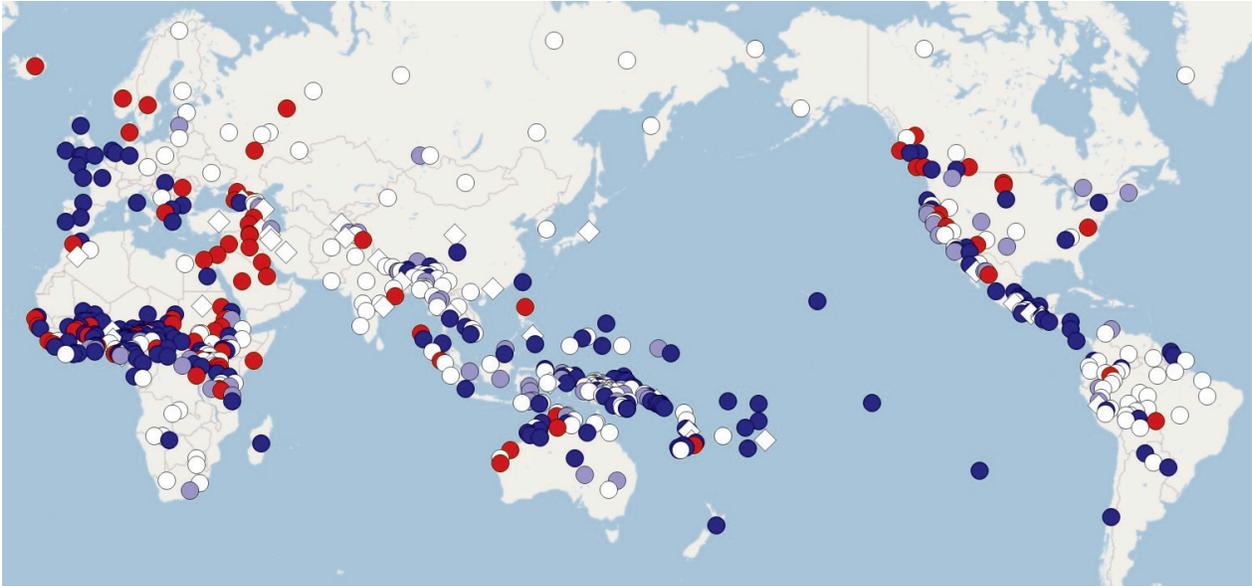


Figure 1.2: Areal distribution of definite and indefinite articles (Dryer 2013b)

in terms of numbers of languages as the other families mentioned above, but since Arawakan languages are mostly spoken in South America, an area with an otherwise relatively low number of articles, I included a higher number of languages from this family due to the relative scarcity of articles in that area of the world.

In certain cases, I include more than one language from the same genus. The genera with more than one languages in my sample are: Kwa (Akan and Logba), as well as Bantu (Limbum and Mokpe), Indo-Aryan (Domari, Palula, and Rajbanshi), Himalayish (Bantawa and Lepcha), Oceanic (Maori, Rapa Nui, Siar Lak, Sunwadia, Sye, and Tongan), Caribbean Arawak (Arawak and Wayuu), and Southern Maipuran (Asheninka Perené and Baure). Oceanic, Bantu, and Indo-Aryan are very large genera, most of them covering a large geographical area; the languages included in the sample from these genera generally have different types of article systems. The second point also holds for the other languages from single genera in the sample; the two Caribbean Arawak languages were included because articles are generally not very common in South America languages.

Another factor that determined the choice of languages was to include as many recent grammars as possible. The methodology for language documentation and grammatical descriptions has made considerable advances over the last decades; recent descriptions often provide more examples and better glosses, leading to more transparency in the primary sources.

In addition to the data from the sample described above, I also discuss primary data from a few languages. Table 1.3 gives an overview of the languages, the number, age, and sex of the consultants, as well as the locations and dates.⁷ From the languages for which I have primary

⁷I thank my colleagues Jude Nformi and Sampson Korsah for their help as consultants and for providing me with data from Limbum and Akan.

data, Mokpe, Limbum, and Akan are also part of the 104-languages sample. Kaqchikel, Q’anjobal, Tz’utujil, and Ejagham are not part of the sample but serve as illustrations in different parts of the study. The languages from the sample are listed in Tables 1.4 to 1.9. The family and genus information is taken from their classification in Glottolog 4.3 (Hammarström et al. 2020).

The presentation of examples from languages other than English follows the Leipzig glossing rules.⁸ All abbreviations used in the glosses are given in the list of abbreviations. To indicate that the presence of a certain marker is obligatory in a given context, it appears in round brackets with an asterisk outside of the brackets: *(marker). To indicate that a certain marker is infelicitous in a given context, it appears in round brackets with an asterisk inside of the brackets: (*marker). Since this study includes various examples of articles from different languages that occur on either sides of the nouns, I use square brackets whenever they are helpful to indicate which noun the article belongs to, e.g. [noun article]. Note that this is only a matter of presentation and not intended to represent a particular syntactic structure. Also, to unify the glosses across various languages and sources, I sometimes deviate from the original glosses, especially from those of the articles in order to allow for consistent labels in the present study.

Table 1.3: Primary data collection

language	location	time	consultants
Mokpe	Buea (Cameroon)	Feb 2017	female, 21 male 26
Ejagham	Buea (Cameroon)	Feb 2017	male, 28
Limbum	Buea (Cameroon) Leipzig (Germany)	Feb 2017 Feb 2018	male, 29
Akan	Leipzig (Germany)	Jun 2017	male, 31
Tikuna	Bogotá (Colombia)	Oct 2017	male, 28 male, 25
Kaqchikel	Patzún (Guatemala)	Nov 2017	female, 36 female, 27 female, 65
Q’anjobal	Santa Eulália (Guatemala)	Nov 2017	male, 65 male, 41
Tz’utujil	San Antonio (Guatemala)	Dec 2017	female, 35 male, 38

⁸Available at <https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>.

Table 1.4: Languages from the macro area “Africa”

language	family / genus	article	reference	original label
	Afro-Asiatic			
Konso	Cushitic	DEF	Orkaydo (2013: 95)	def. reference
		EXSPEC	Orkaydo (2013: 94)	indef. reference
Buwal	Chadic	DEF	Viljoen (2013: 237)	def. determiner
		EXSPEC	Viljoen (2013: 234)	indef. determiner
	Atlantic-Congo			
Akan	Kwa	ANA	Amfo (2010), Arkoh (2011), Arkoh & Matthewson (2013),	(familiar) def. article
	Volta-Congo	EXSPEC	Bombi (2018), Bombi et al. (2019), Sampson Korsah (pc)	(specific) indef. article
Balanta	Central Atlantic	DEF	Fudeman (1999: 310)	def. article
Ganja				
Bullom So	Mel	DEF	Childs (2011: 77)	def. article
Limbang	Bantu (Narrow Grassfields)	ANA	Fransen (1995), Jude Nformi (pc)	demonstrative
Logba	Kwa	DEF	Dorvlo (2008: 88)	def. marker
	Volta-Congo	EXSPEC	Dorvlo (2008: 89)	indef. marker
Mokpe	Sawabantu	DEF	primary data, Atindogbe (2013: 11)	def. article
Oko	Benue-Congo	DEF	Atoyebi (2010: 127)	def. article
		EXSPEC	Atoyebi (2010: 128)	indef. article
Supyire	Senoufo	DEF	R. Carlson (1994: 77)	def. suffix
Bambassi	Blue Nile Mao Creoles	DEF	Ahland (2012: 282)	def. article
Angolar	Creole	EXSPEC	Lorenzino (1998: 132)	indef. article
Kabuverdianu	Creole	INDEF	Baptista (2002: 24)	indef. article
Sheko	Dizoid	DEF	Hellenthal (2010: 138)	def. marker
Jamsay	Dogon	DEF	Heath (2008: 255)	def. particle
Gaahmg	Eastern Jebel	DEF	Stirtz (2011: 134)	def. clitic
Lango	Nilotic	PRES	Noonan (1992: 161)	indef. marker
Beng	Mande	DEF	Paperno (2014: 48)	def. article
Sandawe	Sandawe	DEF	Steeman (2011: 72)	def. marker
Koyra Chiini	Songhay	DEF	Heath (1999: 98)	def. morpheme
Dime	South Omotic	DEF	Seyoum (2008: 42)	def. marker

Table 1.5: Languages from the macro area “Eurasia”

language	family / genus	article	reference	original label
Barwar	Afro-Asiatic Semitic	EXSPEC	Khan (2008: 450)	indef. particle
Kharia	Austroasiatic Mundaic	PRES	Peterson (2011: 142)	no label
Sri Lanka Malay	Austronesian Malayic	INDEF	Nordhoff (2009: 319)	indef. marker
Basque	Basque	REF PRES	Hualde & Ortiz de Urbina (2003: 118-122)	def. article indef. article
Tamil	Dravidian	INDEF	T. Lehmann (1993: 112)	indef. adjectival
Albanian	Indo-European Albanian	DEF INDEF	Newmark et al. (1982: 149-150), Buchholz & Fiedler (1987: 232)	def. article indef. article
Armenian	Armenic	DEF INDEF	Dum-Tragut (2009: 108) Dum-Tragut (2009: 105)	def. article indef. article
Domari	Indo-Aryan	INDEF	Matras (2012: 103)	indef. marker
German	Germanic	DEF ANA DEF _{WEAK} INDEF	Hentschel & Weydt (2003: 208-211), Schwarz (2009)	def. article strong def. art. weak def. art. indef. article
Irish	Celtic	DEF	Stenson (2008: 35)	def. article
Palula	Indo-Aryan	ANA EXSPEC	Liljegren (2016: 140-142)	anaphoric art. indef. article
Rajbanshi	Indo-Aryan	EXSPEC	Wilde (2008: 71)	indef. article
Bonan	Mongolic-Khitani Moghli	INDEF	Fried (2010: 67)	indef. enclitic
Aghul	Nakh-Dagestanian Lezgi	INDEF	Ganenkov et al. (2009: 7)	indef. marker ⁹
Ao (Mongsen)	Sino-Tibetan Kuki-Chin- Naga	ANA	Coupe (2007: 110)	anaphoric demonstrative
Bantawa	Himalayish	ANA	Doornenbal (2009: 96)	def. marker
Hakhun	Brahmaputran	DEF	Boro (2017: 225)	def. article
Tangsa				
Lepcha	Himalayish	DEF	Plaisir (2006: 54)	def. article
Turkish	Turkic Oghuz	INDEF	Göksel & Kerslake (2005: 179)	indef. article
Hungarian	Uralic	DEF INDEF	Kenesei et al. (1998: 256-259)	def. article indef. article

⁹The original Russian label is “показатель неопределенности”.

Table 1.6: Languages from the macro area “Papunesia”

language	family / genus	article	reference	original label
Menya	Angan Austronesian	DEF	Whitehead (2006: 53)	definite clitic
Agta	Northern Luzon	DEF EXSPEC	Robinson (2008: 92) Robinson (2008: 93)	definite marker specificity marker
Bajau (West Coast)	Basap-Greater Barito	DEF	Miller (2007: 116)	demonstrative
Biak	Greater SHWNG	DEF EXSPEC NSPEC	van den Heuvel (2006: 202-218)	givenness marker specific marker nonspec. marker
Maori	Oceanic	INSPEC PRES	Bauer (1993: 350-359), Harlow (2012: 140-145), Polinsky (1992), Chung et al. (1995)	def. article indef. article
Rapa Nui	Oceanic	REF	Kieviet (2017a: 237-241), Weber (2003)	referential article
Siar Lak	Oceanic	INSPEC NSPEC	Rowe (2005: 15-18)	referential article non-ref. article
Sunwadia	Oceanic	REF	Henri (2011: 116-128)	article ¹⁰
Sye	Oceanic	PRES	Crowley (1998: 161)	indef. premodifier
Tongan	Oceanic	ANA INSPEC _{WEAK} NSPEC	Churchward (1985), MacDonald (2014), Hendrick (2005)	def. accent spec. article nonspec. article
Bilua	Bilua	PRES	Obata (2003: 79)	indef. determiner
Urama	Kiwaian	ANA DEF	J. Brown et al. (2016: 23) J. Brown et al. (2016: 21)	anaph. demonstr. def. article
Komnzo	Yam	ANA PRES	Döhler (2018: 110) Döhler (2018: 102)	anaph. demonstr. indef. determiner
Nuclear Trans New Guinea				
Oksapmin	Asmat-Awyu-Ok	RECOG DEF INDEF	Loughnane (2009: 123) Loughnane (2009: 128) Loughnane (2009: 120)	recog. demonstr. def. demonstr. indef. demonstr.
Lavukaleve	Lavukaleve	RECOG DEF INDEF	Terrill (2003: 83) Terrill (2003: 59) Terrill (2003: 78)	specifier def. article indef. marker
Timor-Alor-Pantar				
Bunaq	East Timor-Bunaq	ANA	Schapper (2009: 276)	def. article
Teiwa	Alor-Pantar	ANA PRES	Klamer (2010: 137) Klamer (2010: 212)	demonstrative indef. marker

¹⁰Henri (2011) distinguishes between “les articles des noms communs animés” and “les articles des noms communs inanimés”.

Table 1.7: Languages from the macro area “Australia”

language	family / genus	article	reference	original label
Gooniyandi	Bunaban Gunwinyguan	RECOG	McGregor (1990: 145)	determiner
Bininj Kun-Wok	Marne	RECOG	Evans (2003: 297)	recog. demonstr.
Wubuy	Gunwinyguan Bak	ANA	Heath (1984: 269)	anaph. demonstr.
Mangarrayi	Mangarrayi- Maran	ANA	Merlan (1989: 43)	anaph. prefix
Guragone	Maningrida	ANA	R. Green (1995: 146)	anaph. demonstr.
Jingulu	Mirndi	ANA	Pensalfini (1997: 232)	anaph. demonstr.
Bardi	Nyulnyulan	ANA	Bowern (2012: 323), Bowern (2008: 79)	
Martuthunira	Pama-Nyungan Pilbara	ANA	Dench (1994a: 111, 114)	topic tracking demonstrative
Yankuntjatjara	Desert Nyungic	RECOG	Goddard (1985: 54)	anaph. demonstr.
Yuwaalaraay	Southeastern Pama-Nyungan	ANA	Giacon (2014: 189), Williams (1980: 91)	anaph. demonstr.

Table 1.8: Languages from the macro area “North America”

language	family / genus	article	reference	original label
Yurok	Algic	DEF	Garrett (2010: 23)	article
Ventureño	Chumashan	INDEF	Henry (2012: 166)	indef. article
Haida	Haida	DEF	Enrico (2003: 753)	def. suffix
Chimariko	Chimariko	DEF	Jany (2009: 73)	def. suffix
Ch’ol	Mayan	DEF INDEF	Vázquez Alvarez (2011: 245) Vázquez Alvarez (2011: 246)	def. article indef. determiner
Ulwa	Misumalpan	DEF PRES	T. Green (1999: 285, 157), Koontz-Garboden (2009)	def. determiner indef. article
Zoque (Chimalapa)	Mixe-Zoque	DEF	Johnson (2000: 243)	def. article
	Otomanguean			
Chatino (Teotepec)	Zapotecan	PRES	McIntosh (2011: 83), McIntosh (2015)	numeral, indef. art
Chocholtec	Popolocan- Mazatecan	PRES	Campbell & Luna Villanueva (2011: 21)	det. adjective ¹¹
Sahaptin	Sahaptian	ANA	Jansen (2010: 197)	demonstrative
Halkomelem	Salishan	REF	Suttles (2004: 339)	articles
Lakota	Siouan	DEF EXSPEC NSPEC	Van Valin (1977: 62), Williamson (1984: 47), Rood & Taylor (1996), Ingham (2003: 90)	def. article spec. article nonspec. article
Tepehua (Huehuetla)	Totonacan	REF PRES	Kung (2007: 385) Kung (2007: 385)	def. article indef. article
	Uto-Aztecan			
Cupeño	California Uto-Aztecan	DEF	Hill (2006: 351)	determiner
Pipil	Corachol- Aztecan	DEF INDEF	Campbell (1985: 56) Campbell (1985: 56)	def. article indef. article
Tarahumara	Tarahumara- Guarijio	DEF	Cohen (1998: 160)	def. article
Ute	Numic	DEF	Givón (2011: 163)	def. article
Nuuchahnulth	Wakashan	DEF	Davidson (2002: 297)	def. article

¹¹The original Spanish label in Campbell & Luna Villanueva (2011: 21) is “adjetivo determinativo”.

Table 1.9: Languages from the macro area “South America”

language	family / genus	article	reference	original label
Mapudungun	Araucanian	PRES	Smeets (2008: 81)	indef. article
	Arawakan			
Arawak	Caribbean	DEF	Pet (2011: 43, 52, 88)	def. article
	Arawakan	INDEF		indef. article
Asheninka	Southern	PRES	Mihas (2010: 182)	indef. article
Perené	Maipuran			
Bare	Medio Rio Negro	ANA	Aikhenvald (1995: 23)	anaph. demonstr.
Baure	Southern Maipuran	REF	Danielsen (2007: 312)	article
Parecis	Central Maipuran	PRES	Barros Brandão (2014: 101)	indef. article
Wayuu	Caribbean	DEF	Álvarez (2017: 12)	det. article
	Arawakan	PRES	Álvarez (2017: 49)	indef. article ¹²
Carib	Cariban	INDEF	Courtz (2008: 139)	indef. article
Aguaruna	Chicham	ANA	Overall (2007: 157)	anaph. demonstr.
	Tupian			
Akuntsú	Tuparic	DEF	Aragon (2014: 162)	determinative
Nhengatu	Tupi-Guarani	PRES	da Cruz (2011: 265)	indefinite ¹³
Apinayé	Nuclear-Macro-Je	DEF	Cunha de Oliveira (2005: 165)	def. article
	Nambiquaran			
Mamaindé	Nambikwara Complex	REF	Eberhard (2009: 365)	article
Sabanê	Sabanê	REF	Antunes de Araujo (2004: 92)	referential suffix
	Pano-Tacanan			
Kashibo-Kakataibo	Panoan	PRES	Zariquiey Biondi (2011: 303)	indef. marker
Cubeo	Tucanoan	ANA	Chacon (2012: 331)	anaph. determ.
Ayoreo	Zamucoan	INSPEC	Bertinetto (2009: 23)	base form
		NSPEC	Ciucci (2016: 482)	indeterminate form

¹²The original Spanish labels in Álvarez (2017) are “artículo determinado” and “artículo indefinido”.

¹³The original Portuguese label in da Cruz (2011) is ‘determinante indefinido’.

Chapter 2

Articles as a crosslinguistic category

What is an article and how might articles be described in a maximally general way such that the difference between a language that has these elements and one that does not can be characterized in a meaningful manner? (Moravcsik 1969: 85)

2.1 The problem

Although the question of how articles can be defined for crosslinguistic comparison is not a new one, we still lack a good answer to it. This might be due to the fact that especially in European languages, identifying articles is not a very complicated matter. With no established tradition of what counts as an article and what does not, however, the classification of markers in other languages is often less obvious. A simple example of a well-established article system is shown in (2), illustrating the definite and indefinite articles in English in opposition to each other.

- (2) a. The woman got off **the** car.
b. The woman got off **a** car.

Glossing over the details for now, we can say that the definite article in (2a) marks the referent of *car* as identifiable, known, given, or accessible, and also as prominent in the scenario described, while the indefinite article in (2b) marks it as not uniquely identifiable and also as less prominent.

We find this kind of opposition of adnominal markers that encode identifiability or non-identifiability of discourse referents in many European languages. Although there are differences in the actual distributions of such markers or articles in different languages, the labels and properties of a few well-known article systems are often imposed to other, less described languages. Therefore, it is not surprising that markers of less known languages are often not regarded as articles. This is problematic, since without having a clear picture of the crosslinguistically relevant properties

of articles, we do not know which properties of e.g. the definite article in English are category-defining and which ones are less central, and we might end up classifying markers as articles or not only on the basis of their similarity to the articles found in English.

The following paragraphs present a number of examples that illustrate this problem. Examples (3) and (4) below contrast two functionally similar if not equivalent expressions: in the case of German, we would traditionally label *ein* as an indefinite article; the equivalent Chinese expression consisting of the numeral ‘one’ and a classifier has similar referential functions in this context, but one may not regard *yi zhuang* in (4) as an indefinite article.

- (3) Ich habe letztes Jahr **ein** Haus gekauft.
 I have last year **ART:INDEF** house bought
 ‘I bought a house last year.’
 German (primary data)

- (4) Ta qunian mai le [(yi) **zhuang** fangzi].
 3SG last.year buy PFV **one CL** house
 ‘He bought a house last year.’
 Chinese (P. Chen 2003: 1171)

For the Chinese construction, it is not as clear as for the German one whether or not it involves an indefinite article. In the literature, this construction with *yi* ‘one’ and a classifier has been argued to be used in various contexts corresponding to those of indefinite articles in other languages (cf. P. Chen 2003). It is even attested in predicative contexts which are no longer referential:

- (5) Ta shi [(yi) **ge** maimairen].
 3SG be **one CL** businessman
 ‘She is a businessman.’
 Chinese (P. Chen 2003: 1171)

Similarly, the possessive markers in a number of Uralic languages are difficult to properly classify, as some are distributed in a very similar way to definite articles in other languages. This is what we find in, for instance, Udmurt (cf. Section 4.2.1). Although still used as the third person singular possessive marker, *-ez/-ze* has been shown to indicate definiteness even in contexts with no possessor being semantically present or recoverable (e.g. Gerland 2014, Serdobolskaya et al. 2019).

- (6) nêl-êz
 girl-**POSS:3SG**
 ‘his/her daughter’
 Beserman Udmurt (Serdobolskaya et al. 2019: 291)

- (7) Tože kartoška pun-iš'ko=no, marêṃ zək parš'-ez, so kət-âz ug târ
 also potato put-PRS:1SG=ADD HESIT big pig-POSS:3SG that belly-POSS:3SG NEG fill.SG
 leš'a.
 evidently
 '[The speaker is explaining how she feeds her pigs – with goose-foot, coltsfoot etc.] I also
 give them potatoes, the big pig is not full up.'
 Beserman Udmurt (Serdobolskaya et al. 2019: 305)
- (8) Tolon vel't-i-m č'aš'a-e gibija-nê, so bere č'aš'a
 yesterday go-PST-1PL forest-ILL pick.mushrooms-INF that after forest
 gibi-ze ujbât š'už'ja-m.
 mushroom-POSS:ACC.3SG whole.night clean-PST.1PL
 'We went to the forest to pick mushrooms and then we spent the whole night cleaning
 [the] mushrooms.'
 Beserman Udmurt (Serdobolskaya et al. 2019: 306)

In (6), we see *-âz* in its prototypical function, marking the possessor on the possessum *nâl* 'daughter'. In (7) and (150), on the other hand, *-ez/ze* no longer marks a third person possessor. The context of (7) includes a first person possessor, given that the pig belongs to the speaker of the utterance. In (150), on the other hand, no possessor is semantically present, and *-ze* is used to indicate that the referents of *gibi* 'mushrooms' are identifiable (by means of previous mention) and hence definite. Again, this shows how much functional overlap definite articles can have with other types of markers and how difficult it is to distinguish between them. Whether or not such possessive markers, distributed in a similar way to definite articles in other languages, should be treated as definite articles thus requires clear, language-independent criteria.

Another entirely different phenomenon that touches upon the definition of articles as a crosslinguistic category is usually referred to as "nominal tense". Nominal tense is found in different languages of South America (i.e. Arawakan, Cariban, Movima, Matacoan, Tupi-Guarani) and refers to temporal and/or deictic marking by grammatical elements in the nominal domain. In some cases, nominal tense combines with temporal marking on the verb, and sometimes it is the only grammatical indicator for temporal relations (Campbell 2012: 285). What is relevant for articles and definiteness is that in some languages, these temporal markers on the noun can lead to an either anaphoric or nonspecific interpretation of the referent. Example (9) from Nivaclé (Matacoan, Paraguay) shows how different past markers on the noun mark the referent as anaphoric and thus as definite. In (9a), *-naxi* indicates yesterday's past and *-mati* is used in (9b) to mark today's past.

- (9) a. mansana Ø-tolu Ø-hile-naxi hohnat wit hi-kwes.
 apple 3-come.from POSS:3-tree-PST2 ground CONJ 3-split
 'The apple fell from the tree (that we saw yesterday) and split.'

- b. *sinox-mati atana Ø-yil-li.*
 dog-**PST1** now 3-die-**ITER.SG**
 ‘The dog (from earlier today) is sick now’.
 Nivaclé (Campbell 2012: 287)

Since these two markers also have referential functions to a certain extent by encoding the referent as anaphoric, more concrete criteria for treating a marker as an article or not are needed in order to include or exclude such nominal tense markers.

Those three examples from Chinese, Udmurt, and Nivaclé showed that the properties of articles from well-known European languages may not be sufficient to define articles as a category in a language-independent way. This is especially important if one takes into account other non-referential properties of (European) articles as well. For instance, we find definite articles in generic contexts (which do not involve definiteness) in some languages, but in others, articles are not required in such contexts, leading to variation even within European languages. Examples (10) and (11) show German and English, which do not use a definite article with a generic referent in the plural, while (12) and (13) show that Spanish and Hungarian require it:

- (10) I like [(***the**) cats].
- (11) Ich mag [(***die**) Katzen].
 1SG like.1SG **ART:DEF** cats
 ‘I like cats.’
 German (primary data)
- (12) Me gustan [**(los)** gatos].
 1SG.DAT please.3PL **ART:DEF** cats
 ‘I like cats.’
 Spanish (primary data)
- (13) Szeretem [**(a)** macskákat].
 like.1SG **ART:DEF** cats
 ‘I like cats.’
 Hungarian (primary data)

The examples from this section showed that there are two main issues that need to be addressed when defining articles as a crosslinguistic category. First, we see that even in European languages, the distribution of articles varies, and a distributional criterion for articles must take this into account. Second, the fact that definite articles occur with generic referents shows that the use of articles is not necessarily restricted to their referential values. Taking these two issues into account, the aim of this chapter is to identify the properties of articles which define them as a crosslinguistic category and which allow for a comparison of articles across languages.

2.2 Defining articles

2.2.1 Articles as a crosslinguistic category

A suitable approach to define articles for crosslinguistic comparison is the one of comparative concepts, proposed in Haspelmath (2010).¹ Comparative concepts “are not part of particular language systems and are not needed [...] by speakers. Comparative concepts are universally applicable, and they are defined on the basis of other universally applicable concepts: universal conceptual-semantic concepts, general formal concepts, and other comparative concepts” (Haspelmath 2010: 665). In other words, comparative concepts are crosslinguistic categories that are defined in a way so that they can be applied to any given language, without relying on language-specific criteria. In the present study, I will adopt this approach to define a crosslinguistic category of articles to avoid the risk of using certain language-specific properties of well-described articles to classify lesser described markers in other languages. For instance, only because the definite article in English is a free form and no affix (at least represented as such in the orthography), one should not generalize that articles need to be free forms.

The definition of articles in Dryer (2014: 234) is similar in that respect and counts as a definition of articles as a crosslinguistic category. However, it is formulated in a rather vague way which makes it difficult to be applied consistently across languages.

[...] I use the term ‘article’ in a somewhat nonstandard sense, one that is more semantic than common uses of the term. Specifically, although I restrict it to words or morphemes that occur in noun phrases, I otherwise make no assumptions as to the syntactic status of the words or morphemes in question. Namely, they must code something in the general semantic domain of definiteness or indefiniteness [...]. (Dryer 2014: 234)

Lyons (1999) provides another definition of articles. Importantly, he emphasizes that the function of articles should not be restricted to the marking of (in)definiteness, which is one of the central points of this study as well.

We will see, however, that articles can encode more than definiteness or indefiniteness, and that they have been argued to have a quite different principal function, at least in some languages. (Lyons 1999: 6)

In either cases, articles are defined (or characterized) based on their referential functions, while it remains implicit how the definitions can be applied in practice.

¹The comparability of linguistic categories (mostly syntactic categories, syntactic operations based on syntactic categories, and parts of speech or lexical categories) across languages is not a new topic and was extensively discussed by a number of authors, e.g. Cristofaro (2009), Croft (1991), Culicover (1999), Dryer (1997), Haspelmath (2008), Lazard (2005), Newmeyer (2007), Slobin (1997, 2001), Song (2001), Stassen (1985), Wierzbicka (1998).

In contrast to these two definitions of articles as a single, crosslinguistic category with no further distinction of different types, Schroeder (2006) provides two separate definitions for definite and indefinite articles, respectively. He combines formal and functional criteria for his definitions, the latter being presented in a rather condensed way as well. As I will argue in Chapter 3 following Hawkins (1978), Himmelmann (1997), and Lyons (1999), definiteness and indefiniteness are no primitive semantic or pragmatic categories and call for a more fine-grained distinction in order to define different types of articles.

Definite articles will be defined as a morphological device (free morpheme, enclitic morpheme, or affix) with the primary function to denote the definiteness of noun phrases with anaphoric and/or nonanaphoric reference. [...]

“Indefinite article” will be defined as a morphological device (free morpheme, enclitic morpheme, or affix) with the primary function to denote the indefiniteness of noun phrases. (Schroeder 2006: 553)

The goal of this section is to define articles as a crosslinguistic category based on criteria that can be applied to any marker in any given language, which allows us to discuss and compare articles in the world’s languages. With this in place, I will further define 10 different types of articles on the basis of their referential functions in Chapter 3.

I propose three main criteria to define articles as a crosslinguistic category. The first criterion is their function: articles are defined as referential markers. The second criterion refers to their domain in that articles have to occur together with lexical referential expressions. In most languages, such lexical, referential expressions will be nouns. However, nouns as a morphosyntactic category are not universal and can therefore not be used as a comparative concept. The third criterion is concerned with the distribution of articles and requires them to occur systematically (allowing for exceptions) in the contexts matching their referential functions. The definition of articles is given in (14).

(14) *Articles as a crosslinguistic category*

Articles are markers in the nominal domain whose main function it is to encode the referential function of the lexical referential expression that they occur with.

In the following sections, I will discuss those three criteria in detail with examples from different languages.

2.2.2 Articles encode referential functions

Articles encode the referential functions of the lexical referential expression (noun) that they occur with. In this section, I briefly present a selection of referential functions that need to be

distinguished to define different types of articles (see Sections 3.2 and 3.3 for a more detailed overview). I then present two examples of markers which also encode referentiality but which are not classified as articles here because their primary function is a different one.

We can distinguish three main referential functions: definite, specific, and nonspecific. Definite referents are unambiguously identifiable by the speaker and the hearer; there are different subtypes of definite referents which differ in the way identifiability is established. Deictic definites are identifiable by ostension, i.e. physical perception. This establishes a link between an object that is present in the discourse situation and the referent, as in (15).

(15) Do you see **the house** over there? (deictic)

Anaphoric referents, on the other hand, are fully identifiable by the speaker and the hearer due to previous mentioning in the discourse. In (16b), the referent of *shoes* is used anaphorically and can therefore be marked by the definite article *the*:

(16) a. Do you like **my new shoes**? Do you think I can wear them to the event tonight?
b. Yes, you should wear **the shoes** shoes tonight. (anaphoric)

Another type of definite referents are recognitional definites. Such referents are unambiguously identifiable by shared experience between the speaker and the hearer, or by common (personal or cultural) knowledge. This allows for the first mention of *the dog* in (17) to be marked with the definite article.

(17) What happened to **the dog** (we used to have)? (recognitional)

Unique referents represent another type of definite referents because they are the only (salient) referent of their kind. This can hold for the discourse situation as in (18), in which the referent of *kitchen* is unique in the immediate discourse situation. In (19), on the other hand, the referent of *president* is unique in a larger context than the immediate discourse situation; using the definite article constructs the referent as unique in this case.

(18) Where is **the kitchen** in your flat? (situationally unique definite)

(19) **The president** gave a speech. (contextually unique definite)

In the indefinite domain, we can distinguish between specific and nonspecific referents. Specific referents are no longer unambiguously identifiable by both the speaker and the hearer. However, either of them know that the expression evokes a particular referent of its kind.² Nonspecificity,

²We can further distinguish two types of specific referents: referents that are only non-identifiable for the hearer but not for the speaker, and referents that are non-identifiable for either discourse participant. Since this distinction does not seem to be relevant for articles, I ignore it in this study.

in contrast to definiteness and specificity, does not involve a particular referent evoked by the nominal expression. Any referent of a given kind, existing or non-existing, can be linked to the expression used. The restriction with respect to nonspecific referents is that there can only be one referent linked to the expression, without any restriction on the choice of the referent that is linked. In other words, the discourse situation does not provide enough information to decide which of all the referents of a kind is evoked by the nominal expression. We typically find such referents in interrogative, negative, irrealis contexts, and with intensional verbs. A classic example of the distinction between specific and nonspecific referents is provided in (20), taken from Givón (1984: 441) and going back to Quine (1953):

- (20) John wanted to marry **a rich woman** ...
- a. ...but she refused him. (specific)
 - b. ...but he couldn't find any. (nonspecific)

As English does not obligatorily mark the difference between specific and nonspecific referents on the expression itself, the referent *a woman*, marked by the indefinite article, remains ambiguous between a specific and nonspecific interpretation without further context. Such a contextual disambiguation is provided in the continuations in (20a) and (20b). In (20a), the referent is specific; even though the hearer might not be able to unambiguously identify the referent linked to *a rich woman*, the only valid interpretation is that there is a particular referent of a rich woman that the speaker refers to. In (20b), on the other hand, the referent is marked as nonspecific, since there is no particular referent linked to the expression *a rich woman*. Any referent, existing or nonexistent, that fulfills the criteria of being a *rich woman* can be linked to the expression used, but the context implies that neither the speaker nor the hearer have more information about which referent to link.

As will be shown in Chapter 6, indefinite articles that mark both specific and nonspecific referents are also found in other parts of the world. We do, however, also find languages in which the distinction between a specific and a nonspecific referent is expressed by an article. One example can be found in Bemba (Bantu, Zambia), which has an inclusive-specific article (cf. Section 7.1.1). The inclusive-specific article encodes definite as well as specific indefinite referents, but not nonspecific referents. As (21a) shows for Bemba, the referent of 'book' is necessarily interpreted as definite or specific when it is marked by *i-* as in *icitabo*. When the referent is expressed by the noun without this marker, i.e. as *citabo* in (21b), it must be interpreted as nonspecific.

- (21) a. n-dee-fwaaya i-ci-tabo.
 1SG-PRS-want ART:INSPEC.CL7-CL7-book
 'I want the / a certain book.' (specific)

- b. n-dee-fwaaya ci-tabo.
 1SG-PRS-want CL7-book
 ‘I want a (any kind of) book.’ (nonspecific)
 Bemba (Givón 1984: 65)

Examples (563) and (565) below show for nouns of other classes that including the article necessarily results in a definite or specific indefinite interpretation of the referent. Thus, the augment in Bemba, i.e. the prefix that precedes the noun class marker, can be considered as an article, given that its main function is to indicate the referential status of the noun it occurs with.

- (22) u-mu-ntu aaliishile.
 ART:INSPEC.CL1-CL1-man come.PST.3SG
 ‘The/a man came.’
 Bemba (Givón 1969: 47)

- (23) naalimweene *(a)-ba-ana.
 see.PST.1SG ART:INSPEC.CL2-CL2-child
 ‘I saw (some/the) children.’
 Bemba (Givón 1969: 42)

In some cases, adnominal markers may have referential functions, but their main function is a different one. Such markers are not considered as articles.

One example of such markers can be found in Ughele (Austronesian, Solomon Islands). Ughele has adnominal markers that qualify as articles based on their domain of occurrence. They are clearly determiners that stand in a paradigmatic relation to each other. Frostad (2012: 6) distinguishes between three article types: the personal article *e* which occurs with proper nouns referring to a person (Frostad 2012: 89), the common article *na*, and the focal article *ai*. The two markers *e* and *na*, labelled personal and common article, are shown in (24) and (25):

- (24) meke naghe [e Peni] ...
 and say PROP Peni ...
 ‘and Peni said ...’
 Ughele (Frostad 2012: 100)

- (25) Zioni [na viu] le-lea-na.
 Zioni COMM bird RED-good-ATTR:3SG
 ‘Zioni is a very nice bird.’
 Ughele (Frostad 2012: 97)

Since these two markers occur with different types of nouns, it seems plausible to call them articles. However, their function is different: *e* only occurs with proper nouns referring to persons as in

(24) and *na* is used only with common nouns as in (25). This suggests that their function is not linked to referentiality, but that they mark other lexical properties of the referent (human vs. non-human) besides their syntactic function as a determiner. Therefore, I do not treat these markers as articles.

A different type of marker that involves referentiality and occurs adnominally is found in Koasati (Muskogean, USA). It encodes an absent or deceased referent, which also makes the referent specific, necessarily involving a particular referent of its kind. Examples (26) and (27) illustrate this:

- (26) *yali am-aw-ó:to-k cimpónc-o:to-k ónti-toV.*
 here POSS:1SG-grandfather-EXIST.PST-s Jim Poncho-EXIST.PST-s come.SG-PST
 ‘Here my grandfather Jim Poncho (who is now deceased) came.’
 Koasati (Kimball 1985: 349)

- (27) ... *waciná nahilk-ó:to-n ca-sobáy-ko-k.*
 ... English language-EXIST.PST-O 1SG.STAT-know-NEG.3-SS
 ‘... at that time the English language was unknown to me.’
 Koasati (Kimball 1985: 349)

In (27) we can see that the lexical meaning of “deceased” is extended to mark a former state of the referent, ‘the English language’. The marker *ó:to* is treated as an article in the grammar for language-specific reasons.³ One could argue that it is an article because it encodes specific referents. However, since this seems to be an effect of the lexical meaning of that marker “deceased, no longer present”, I do not regard *ó:to* as a specific article.

2.2.3 Articles occur in the nominal domain

The second criterion for articles restricts their domain; they have to occur in the nominal domain together with nouns or lexical referential expressions. I use “nominal expression” and “noun phrase” (NP) interchangeably in this study, regardless of framework-specific nuances.

The position and form of the article in the noun phrase can be described by partially orthogonal parameters. It can be a separate or affixed morpheme, as the separate article in English, e.g. *the cat*, or the suffixal definite article in Armenian, e.g. *katu-n* ‘the cat’. If the article is an affix, we can further distinguish different types. Affixed articles are often suffixes, but we also find examples of prefixed articles. In Sumu (Misumalpan, Nicaragua), the article affix appears as an infix with

³The marker is part of a paradigm of other adnominal markers that involve the marking of referential functions. Most of these markers are not real articles either, since they are participial forms indicating temporal relations, the referential interpretations rather being an additional effect (Kimball 1985: 344). These markers are probably called articles in the grammar based on their occurrence together with nouns.

a number of nouns that used to be compound forms. An example is *waki-ni-sa* ‘the banana’, in contrast to the form without the article *wakisa* ‘(a) banana’ (Norwood 1997: 22).

Although there might be sufficient language-dependent reasons to classify an article as an affix, a clitic, or a free word form, such a distinction can hardly be motivated in a comparative study. Therefore, I do not aim at a distinction between affixed, cliticized, or free articles here. I propose a distinction between articles that are separable from nouns in the presence of another nominal element, e.g. an adjective, and articles that are not separable from nouns in such contexts. Thus, the article can either stay in its original position adjacent to the noun (“N-anchored”), or surface in a different position at the edge of the entire noun phrase (“NP-anchored”) in the presence of other nominal attributes. This distinction is more comparable across languages and more workable than the distinction between free forms, clitics, and affixes.

Another parameter concerns the linear order. Articles can precede or follow the noun, e.g. German *das Haus* vs. Icelandic *hús-ið* ‘the house’. Paring these two parameters, examples (28) to (31) give an overview of the possible orders of the noun and the article in contexts including adjectives to distinguish between N-anchored and NP-anchored articles:

- (28) *büyük bir oda*
 large ART:INDEF room
 ‘a large room’ (postposed and NP-anchored)
 Turkish (Göksel & Kerslake 2005: 145)
- (29) *juu maqaniyaa lapanak*
 ART:DEF ancient people
 ‘the old people’ (preposed and NP-anchored)
 Tepehua (Kung 2007: 388)
- (30) *kil ticé tilòliyén*
 houses ART:DEF low
 ‘the low houses’ (postposed and N-anchored)
 Bullom So (Childs 2011: 71)
- (31) *kava-yi ságharumɛ ’uway*
 horse-O white.O ART:DEF.O
 ‘the white horse’ (postposed and NP-anchored)
 Ute (Givón 2011: 195)

As these examples show, the classification of articles into N-anchored and NP-anchored ones is not entirely unrelated to the classification of articles into free forms, clitics, and affixes. Articles that are treated as affixes are generally anchored with respect to the noun, also in the presence of other elements in the NP. Articles are often classified as clitics if they are able to appear on

different elements in the NP. Usually, they occur in a fixed position with respect to the entire NP (mostly at the edge), which corresponds to NP-anchored articles. What is traditionally viewed as a free marker, however, also often falls into the category of NP-anchored for staying at the edge of the entire NP.

We do also find languages with a flexible order of the article and the noun, which is the case in Sri Lanka Malay (Malayo-Sumbawan, Sri Lanka, cf. Section 6.1.2). Although described as free variation in Nordhoff (2009: 319-320), the position of the indefinite article may be conditioned by the presence or absence of other elements that occur in the NP, as e.g. the additive marker in (32b).

- (32) a. se **hatthu**=aade.
 1SG ART:INDEF=younger.sibling
 ‘I am a younger sibling’
- b. giini criitha=**hatthu**=le aada.
 like.this story=ART:INDEF=ADDIT exist
 ‘There is also a story like that.’
- Sri Lanka Malay (Nordhoff 2009: 319)

Another formal type of articles can be distinguished if we consider whether or not they are segmental morphemes. Although articles with a nonsegmental exponent are rare in the languages of the world, they are attested. For instance, in Mokpe (Bantu, Cameroon), the definite article corresponds to a tonal process on nouns of certain noun classes. If the noun occurs with a definite article, its first tone-bearing unit appears with a high tone that overrides lexical low tones, observable in indefinite contexts and in the citation form (cf. Section 5.1.2 on the definite article in Mokpe). This is shown in (33) for a number of nouns in isolation.

- (33) a. *mòléli* ‘food’ vs. *móléli* ‘the food’
 b. *líkàlà* ‘(a) bridge’ vs. *líkàlà* ‘the bridge’
 c. *βétàngùlè* ‘(some) lizards’ vs. *βétàngùlè* ‘the lizards’
 Mokpe (primary data)

In the following paragraphs, I present some markers which are article candidates based on their referential function, but which do not meet the domain criterion. Nyaturu (Bantu, Tanzania) has a referential marker that indicates definite referents, but it surfaces in the verbal complex. In contrast to (34a), the agreement marker in the verbal complex in (34b) triggers a definite interpretation of the referent of the object.

- (34) a. n-a-onaa mwalimu.
 s:1SG-PST1-see CL1.teacher
 ‘I saw a teacher.’ Nyaturu (Hualde 1989: 182)

- b. n-a-m σ -onaa mwalimu.
 s:1SG-PST1-O:CL1-see CL1.teacher
 ‘I saw the teacher.’
 Nyaturu (Hualde 1989: 182)

Since there is no additional marker in the noun phrase to indicate the referentiality of the nominal expression, the function of the object marker is comparable to a definite article. In accordance to the intuition that this marker in Nyaturu should not be treated as an article, it can be excluded based on the domain criterion here.

Besides referential markers in the verbal domain, we also find referential markers in the noun phrase that do not qualify as articles because they do not primarily occur with simple nouns and require another element to be present in the noun phrase as well. For instance, in Khwarshi (Tsezic, Russia), we find a definiteness marker that only combines with adjectives and similar expressions.⁴ Accordingly, the marker *so*, shown in (35) and (36), is not treated as a definite article.

- (35) žik’we [miq’e-so baydan] m-ež-i.
 man.OBL.ERG far.away-DEF field.CL:III CL:III-plant-PST.W
 ‘The man planted the furthest field.’
 Khwarshi (Khalilova 2009: 133)

- (36) σ -o λ o λ ’o-so-ho y-o λ o λ ’o-so y-ez-un.
 CL:I-in.middle-DEF-APUD CL:II-in.middle-DEF CL:II-take-PST.UW
 ‘The middle (brother) married the (other) middle (sister).’
 Khwarshi (Khalilova 2009: 133)

While the marker *so* in Khwarshi seems to have other non-referential functions as well, Colloquial Slovenian features a similar marker that is treated as a definite article in Marušič & Žaucer (2014).⁵ Example (37) illustrates the use of this marker *ta*.

- (37) en od ta hitrih avtov
 one.NOM of DEF fast.GEN cars.GEN
 ‘one of the fast cars’
 Colloquial Slovenian (Marušič & Žaucer 2014: 193)

⁴This may appear to be similar to the “strong” and “weak” agreement patterns of adjectives in German, which are sensitive to the presence of inflectional marking on the determiner. This leads to a situation in which different agreement markers are used in combination in the presence of the definite and indefinite article. Diachronically, this difference in agreement markers on the adjective is likely to have had a semantic origin, but from a synchronic perspective, the different agreement patterns on adjectives are no longer conditioned by referentiality itself in German.

⁵Colloquial Slovenian has almost completely lost the old distinction between short and long adjective forms that used to mark the definiteness of the nominal referent. In the standard language, however, we do still find this difference, e.g. *dober pesnik* ‘a good poet’ vs. *dobri pesnik* ‘the good poet’.

However, *ta* can only occur together with attributive adjectives in the noun phrase as in (38a); it cannot mark the referent as definite if no adjective is present (38b). Therefore, I do not regard *ta* as a definite article.

- (38) a. **ta** velika knjiga
 DEF big book
 ‘the big book’
- b. ***ta** knjiga
 DEF book
 intended meaning: ‘the book’
 Colloquial Slovenian (Marušič & Žaucer 2014: 185)

A common assumption based on articles in many European languages is that articles cannot co-occur with other articles within languages, as it is the case in English:

- (39) a. *the a boy
 b. *a the boy

This is also sometimes taken as a criterion against classifying a marker as an article. In this vein, Terrill (2003: 81) argues for a marker in Lavukaleve (Lavukaleve, Solomon Islands) that “[...] *ro* is a marker of indefiniteness, but one would not want to call it an indefinite article, for a number of reasons. [...] Fourthly, and crucially, it can, under certain constrained circumstances, co-occur with the definite article.” In the remainder of this section, I briefly show that we do however find co-occurring articles in the languages of the world, and that this should not be regarded as a criterion for a given marker to be treated as an article or not.

Given that articles are defined as primarily referential markers, it does not seem surprising from a functional perspective that different articles generally do not co-occur. However, the use of articles is rarely exclusively conditioned by their referential function, and we find co-occurring articles in various languages for a number of different reasons. Heine (1997: 70-77) reports that partitives are often expressed by the combination of definite and indefinite articles. Most examples provided, however, involve a partitive marker and not an indefinite article that combines with the definite article, or they show both articles that co-occurred in contexts without nouns. A less controversial example of partitives expressed by the combination of two articles is found in Akan (Amfo 2010). Example (40) below shows the noun *nkɔrɔfo* ‘people’ with both an anaphoric and an exclusive-specific article, and we can see in the translation that this combination has a partitive interpretation.

- (40) [[*nkɔrɔfo no*] *bí*] *ka-a* *sɛ* *wɔ-re-m-pene*.
 people ART:ANA ART:EXSPEC say-COMPL COMP 3PL-PROG-NEG-agree
 ‘Some of the people said that they will not agree.’
 Akan (Amfo 2010: 1795)

The two articles in Akan can also be combined in the reverse order and scope relation; in that case, the noun is indicated as having a non-identifiable specific referent, but the additional anaphoric article indicates that the referent has been mentioned before in the discourse. Example (41) illustrates this:

- (41) [[maame **bi**] **no**] a ɔ-bɛ-hwehwɛ-ɛ wo no ...
 woman **ART:EXSPEC ART:ANA REL** 3SG-MP-look.RED-COMPL you CD ...
 ‘That (certain) woman who came looking for you ...’
 Akan (Amfo 2010: 1797)

Another type of article combinations is the combination of two articles for other, non-semantic motivations. That means that only one of the articles is interpreted semantically, while the other no longer contributes to the interpretation of the referent. We find such a combination of a definite and an indefinite article in Albanian, involving “conflicting” requirements with respect to the use of articles that result in the presence of both articles. The definite article in Albanian is suffixed to the noun, while the indefinite article consists of the free form *një* preceding the noun. In most contexts, the indefinite article cannot combine with a noun that is at the same time marked with the definite article. Only if an indefinite article combines with an adnominal possessive, the noun must also contain the definite article because the possessive marker formally requires the latter to be present. Example (42) shows a noun with a definite referent being marked by a possessive. The noun *libri* ‘the book’ contains the definite article. As example (43) shows, even if the indefinite article is used to mark the referent of the noun as non-identifiable, the definite article is used as well.

- (42) ky është libr-i im.
 this is book-**ART:DEF POSS:1SG**
 ‘This is my book.’
 Albanian (Zymeri 2004: 58)
- (43) një mik-u im
ART.INDEF friend-**ART:DEF POSS:1SG**
 ‘a friend of mine’
 Albanian (Newmark et al. 1982: 150)

2.2.4 Articles are systematically distributed

The aim of this section is to propose a crosslinguistically applicable criterion with respect to the use and the distribution of articles based on what I will call systematicity. Often, obligatoriness is evoked as a criterion for articles, also with the reference to English or “European” articles:

On the basis of what we have seen in English, we may expect articles more widely to act as default members of larger categories of definite or indefinite expressions, to be obligatory (except perhaps under certain generally specifiable conditions) in the absence of other such expressions, and to be unstressed and perhaps phonologically weak. (Lyons 1999: 47)

The crucial point here is that obligatoriness, by default, seems to be understood in a way that the article has to occur in all contexts in which it is expected according to its referential function. However, also in other languages we often find that the use of articles is restricted in many contexts in which they should mark the referential function of the noun (e.g. with mass nouns, plural nouns, nominal expressions with certain adpositions or case markers). As mentioned above, it is problematic to characterize the distribution of articles as obligatory if obligatoriness cannot allow for exceptions.⁶

I will illustrate this with what I treat as an anaphoric article in Hausa (Chadic, Nigeria, cf. Section 5.3.2). Hausa has a marker *-n*, shown in example (44), which is an anaphoric article with extended functions. Example (45) shows a noun which is marked by *-n* and whose referent is definite because of the following relative clause.

- (44) a. *bákâ-n* ‘the bow’
 b. *manòmâ-n* ‘the farmer’
 Hausa (Newman 2000: 144)

- (45) *yārò-n* *dà ya* *tàfi*
 boy-ART:ANA/DEF REL 3SG.M.PERF.REL leave
 ‘the boy who left’
 Hausa (Newman 2000: 145)

In his grammar, Newman (2000: 143) mentions that the label “definite article” is problematic, and that “previous reference marker” would be a better label, indicating that we are dealing with an anaphoric marker that has started to extend its functional domain to other definite contexts. The argument used against treating this marker as a definite article (e.g. in Newman 2000: 143, Lyons 1999: 52, Zimmermann 2008: 419) always is that a bare noun can still be interpreted as definite:

- (46) *tūlū yā* *fashè.*
 pot 3SG.PERF break
 ‘The / a water pot broke.’
 Hausa (Newman 2000: 143)

⁶If obligatoriness is defined in such a way that it can allow for systematic exceptions, one could of course also relate the distributional criterion of articles to obligatoriness. The important point is that one has to allow for systematic exceptions because, to the best of my knowledge, we do not find articles without exceptions in their distribution. I chose the notion of “systematicity” over “obligatoriness” here simply because the latter may more easily hide the fact that factors other than their referential function may condition the distribution of articles.

In example (46), *tùlū* ‘pot’ is not necessarily interpreted as indefinite even in the absence of *-n*. Hence, the presence of the definiteness marker *-n* is not obligatory to mark the noun as definite, so the argument goes.

What the concept of obligatoriness entails is that there is no room for exceptions. However, the behaviour of articles in many languages is always influenced by factors other than referentiality as well; the distribution of articles is never perfect in that sense and always interacts with other parts of the grammar or extra-linguistic factors, as I will show below. The concept of systematicity only requires articles to be distributed systematically in language use, which is not evaluated on the basis of single examples, but on the overall distribution of the marker in the language. This automatically allows for an article to be absent in certain contexts in which its presence would be expected based on its referential function, because other, e.g. syntactic, factors may block the use of the article. The same holds for articles that can appear in contexts in which they are not required as a referential marker, but in which they occur nevertheless, for instance, definite articles that co-occur with demonstratives in certain languages.

In the following paragraphs, I discuss a number of markers that should be treated as articles although they are systematically absent in contexts in which their presence is expected from a referential point of view.

A simple and uncontroversial example of a morphosyntactic restriction concerns the indefinite article in German. It does not have a plural form and cannot combine with nouns in the plural. Example (47a) shows *Katze* ‘cat’ marked as non-identifiable by the indefinite article. Its plural counterpart in (47b) is expressed as a bare plural noun *Katzen* ‘cats’ because the indefinite article is not available in this syntactic context.

- (47) a. Ich sehe [***(eine)** Katze] auf dem Dach.
 I see **ART:INDEF** cat.SG on ART.DEF roof
 ‘I see a cat on the roof.’
- b. Ich sehe **Katzen** auf dem Dach.
 I see **cat.PL** on ART.DEF roof
 ‘I see cats on the roof.’
- German (primary data)

The relevant argument to treat this marker in German as an indefinite article here is that even though nouns can be interpreted as non-identifiable in the absence of the marker, its distribution is systematic and the contexts in which it does not occur can be predicted. Thus, systematically distributed means that exceptions are allowed as long as they can be formulated as rules or strong correlations.

This argumentation may seem trivial with respect to German; however, in less studied languages, such distributional gaps are often pointed out to show that the marker in question cannot

be treated as an article. One example is the “definiteness marker” in Sheko (Hellenthal 2010: 153), which signals the definiteness of the referent as in (48), but does not occur with plural nouns, shown in (49):

- (48) é-ká fáád-k’à-bààb ìn-f yīs-ərá k’òdām-ìn-s yīs
 there-LOC body-in-father **wood-ART:DEF.M** DIST.M-ACC.CONT qodama-ART:DEF-M DIST.M
 há=gīīšū-kōb-t=á kēs-ā-m-ə.
 3SG.M=pull-take-ss=3SG.M go.out.CAUS-put-IRR-STI
 ‘That *qodama* pulls out the wood which is there in the body.’
 Sheko (Hellenthal 2010: 141)

- (49) ak̀n sàm-àb fyāānū-s-k̀n zéénf kī-ít-ə ge-tə ...
 here remain-REL **frog-PL-DAT** well exist-ADDR:PL-STI say-SS ...
 ‘(She) said “stay well!” to the frogs who remained there ...’
 Sheko (Hellenthal 2010: 163)

Since the marker *-f* is only available for singular nouns, Hellenthal (2010: 136) analyses it as being complementary to number marking and does not regard it as an article. However, like the indefinite article in German, it fulfills the distributional criterion which is why I treat it as an article in this study.

Another type of systematic restriction on the use of the article can be found in Mokpe (Bantu, Cameroon). In Mokpe, the definite article has different exponents depending on the class of the noun. Relevant here is the marker *è*, surfacing with nouns of class 1. This marker, as example (50a) shows in opposition to (50b), does not occur with nouns whose first segment is a vowel. Since this restriction of the definite article can be formulated as a simple rule ($\text{è} \rightarrow \emptyset / _V$), it does not pose a problem for treating the marker as an article.

- (50) a. *è paŋga* ‘the sand’
 b. \emptyset *ikpá* ‘(the) salt’
 Mokpe (primary data)

Articles can also be restricted in a systematic way by processes outside of phonology or morphosyntax, e.g. by semantic effects. In many languages, articles do not combine with proper nouns or mass nouns. Example (51a) shows this for the definite article in English which does not combine with (unmodified) proper nouns, while example (51b) illustrates the incompatibility of the English indefinite article with mass nouns.

- (51) a. (***The**) Steve baked (***the**) Hannah a cake.
 b. She is drinking (***a**) milk.

We also find languages with stylistic restrictions on articles. For instance, in Konso (East-Cushitic, Ethiopia), the definite article can be omitted in certain anaphoric contexts according to the grammar (Orkaydo 2013: 97), while it is generally used in anaphoric and other definite contexts (Orkaydo 2013: 95-97). Example (52) illustrates the use of the definite article for referents that have already been mentioned in the discourse:

- (52) *kaasa-sit-n=in* *karmaa-siʔ* *?iff-ay*.
 gun-ART:DEF-INSTR=1 lion-ART:DEF kill-PFV.3.M
 ‘I killed the lion with the gun.’
 Konso (Orkaydo 2013: 97)

In the same story about the lion, the latter does not need to be marked by the definite article every time it is mentioned. Example (53) shows an utterance that comes shortly after the one in (52); the referent of *karmaa* ‘lion’ is definite, but the definite article is not used in this instance.

- (53) *karmaa ka* *ɣapaleeta-asiʔ* *ɣaraa kaassuma=i kaassad-ay*.
 lion and monkey-DEM on question=3 ask-PFV.3.M
 ‘And the lion asked this monkey a question.’
 Konso (Orkaydo 2013: 97)

Since Orkaydo (2013: 97) argues that this is only allowed if the definite referent is highly prominent in the discourse, we can speak of systematic absence of the marker *-siʔ*, which is why I treat it as an article.

Another interesting pragmatic restriction on articles is found in Crow (Siouan, USA). In the so-called elevated narrative genre, specific and nonspecific articles are replaced by a single indefinite marker (Graczyk 2007: 233). Examples (54) and (55) show the “standard” specific and nonspecific indefinite articles *-m* and *-eem*, respectively.

- (54) *dakáak-kaata-m* *húu-laa híi-k*.
 bird-DIM-ART:SPEC come-SS reach-DECL
 ‘A bird came, it reached him.’
 Crow (Graczyk 2007: 228)

- (55) *axée baláxxiikaashe dúup-eem* *alúutkaashe áappaa dí-a-wa-ku-heeʔ*
 father bow two-ART:NSPEC arrow with make-CONT-1-give-AFF
 ‘Father, will you please make me two bows as well as (some) arrows?’
 Crow (Graczyk 2007: 230)

In certain narrative contexts, these markers are not used and are replaced by *-dak/-lak* (*-dak* follows consonants while *-lak* is used after vowels). Note that in this genre, the difference between specific and nonspecific referents is no longer marked, as (56) and (57) illustrate. Example (56)

features a specific referent (*bachee-lák* ‘a man’), a referent that is not identifiable but unambiguously linked to the noun. Example (57) shows the noun *iisashpít-dak* ‘a rabbit’ with a nonspecific referent, i.e. not a particular but a single referent of the kind ‘rabbit’ that is linked to the noun.

- (56) *bachee-lák* *baa-aash-dée-k*.
 man-ART:INDEF INDEF-hunt-go-DECL
 ‘A man went hunting.’
 Crow (Graczyk 2007: 230)

- (57) *éehk bal-héeele-n* *iisashpít-dak* *baappeé-k* *b-aliat-bee-m* *isáa-kaashi-k*.
 that wood-among-LOC rabbit-ART:INDEF 1A.kill-DECL 1A-think-1A-DS large-AUGM-DECL
 ‘There in the woods I thought I killed a rabbit, but to my surprise, it was something very large.’
 Crow (Graczyk 2007: 231)

The concept of systematicity not only allows for the systematic lack of articles in contexts in which they are expected to be present based on their referential functions, but it also accounts for articles that occur in other contexts. An example of articles that are used for reasons other than marking referentiality is the indefinite article in Coptic. It combines with infinitives to form a different non-verbal expression. Together with the marker *hn-* ‘in’, the indefinite article *ou-* derives adverbial expressions from infinitives:

- (58) a. *hn-ou-orch*
 in-ART:INDEF-become.secure
 ‘securely’
 b. *hn-ou-ouōnhebol*
 in-ART:INDEF-show.forth
 ‘openly’
 Coptic, Sahidic dialect (Plumley 1948: 40)

Another syntactic use of articles can be illustrated with indefinite articles from Q’anjob’al (Mayan, Guatemala). The language has an exclusive-specific article and classifiers, but no definite article.⁷ The exclusive-specific article as well as classifiers precede the noun. In order to use an adnominal demonstrative which follows the noun, Q’anjob’al requires this prenominal position to be filled. The default marker filling that position is the classifier, as is illustrated in examples (59) and (60).

- (59) **cham** *icham tu’*
 CL:MAN man DEM:DIST
 ‘that man’
 Q’anjob’al (primary data)

⁷The classifiers do not have (primary) referential functions.

- (60) te na tu'
 CL:TREE house DEM:DIST
 'that house'
 Q'anjob'al (primary data)

If no classifier is available for the noun in question, the indefinite article is used instead. Examples (61) to (63) show the use of the exclusive-specific article together with the demonstrative.

- (61) jun ora tu'
 ART:EXSPEC hour DEM:DIST
 'at that hour'
 Q'anjob'al (primary data)
- (62) jun b'eqan ti'
 ART:EXSPEC moment DEM:PROX
 'at this moment'
 Q'anjob'al (primary data)
- (63) jun tz'uy ti'
 ART:EXSPEC bag DEM:PROX
 'this bag'
 Q'anjob'al (primary data)

Without going into detail with regard to the distribution of the exclusive-specific article and classifiers, the important point is that the noun does not combine with the demonstrative without any element filling the prenominal position. The default marker being the classifier, the exclusive-specific article can also take over this function without marking referentiality.

The last example I discuss in this section shows a marker that qualifies as a definite article but cannot be treated as such due to its unsystematic distribution.⁸ Tz'utujil (Mayan, Guatemala) has a marker *ja(la)*, which can occur in different definite contexts. Example (64) shows *ja* in an anaphoric context; in (65), we see marking a referent as situationally unique.

- (64) a. taq xinipoon p-jaay xintzu' [jun ixoq] k'in ik'ee' ak'ala'
 when arrive.PST.1SG LOC-house see.PST.1SG ART:INDEF woman with two children
 etz'b'ula' chuchii' jaay
 seated in.front.of house
 'When I came home, I saw a woman with two children sitting on my doorstep.'

⁸It is probably more accurate to say that the factors that condition the use of *ja* in Tz'utujil are unclear rather than its use being systematic. Systematicity is meant to apply to the referential distinctions; it rather means that the use of the marker is constrained by additional factors which lead to an unsystematic or irregular use with definite referents.

b. awtaqiin ruwach [ja ixoq]?
 know.PST.2SG eyes DEF woman
 ‘Did you know that woman?’
 Tz’utujil (primary data)

(65) b’anitzra k’ool wa [ja q’atb’al] tziijchpam jawa tinamet?
 where EXIST Q DEF town.hall inside DEM town
 ‘Where is the town hall in this town?’
 Tz’utujil (primary data)

In addition, *ja* does not occur in indefinite contexts, as is illustrated for specific referents in (66) and for nonspecific referents in (67). In both of the contexts, Tz’utujil uses the indefinite article *jun* instead.

(66) k’oola [jun sepaneem] awyin tat
 EXIS.1SG ART:INDEF gift for.2SG you
 ‘I have a gift for you.’
 Tz’utujil (primary data)

(67) k’ool [jun atz’ib’ab’el]?
 EXIST ART:INDEF pen
 ‘Do you have a pen?’
 Tz’utujil (primary data)

Treating *ja* as a definite article in Tz’utujil is problematic because its use differs to a great extent among speakers. According to my consultants, age and dialectal variation may influence the use of this marker as a definite article. However, also single speakers use this marker very inconsistently. Often, both a bare noun and a noun marked by *ja* can occur in definite contexts, and according to my consultants, there seems to be neither a difference in meaning nor a preference for one version or the other. Perhaps the marker *ja* is used systematically and we only do not understand at this point the factors that determine its use. However, based on the situation as presented above, I do not treat *ja* as a definite article, even though it appears to mark the referent of the noun it occurs with as definite. The fact that speakers have no clear preference in favor or against the use of *ja* with definite referents, I cautiously conclude at this point that *ja* in Tz’utujil may not be distributed in a systematic way, at least not as a definite article.⁹ Similar examples of markers whose use in definite contexts is rather a by-product from pragmatic factors that condition their use with discourse-prominent referents will be discussed in Sections 4.2 and 5.3.

⁹In the grammar of the Guatemalan Academy of Mayan languages, *ja* is treated as a definite article without further discussion regarding its distribution (Tz’utujil Tinaamitaal 2007: 87-90). Also Dayley (1985: 254-256) treats *ja* as a definite article, providing some contexts in which the marker can be used, noting that it also often occurs with pronouns in topical positions. Thus, *ja* may not be an article but a topic marker which is highly correlated with definite referents.

This last example of the marker *ja* in Tz'utujil shows the limitations of the current approach. Since the present study mostly relies on data from reference grammars and language descriptions or primary data collected using questionnaire, it is difficult to apply more rigorous, quantitative criteria as to what counts as systematic and what does not. Ideally, systematic occurrence should be tested quantitatively with corpus data; for instance, a number of different types of referential contexts could be identified and tested for the occurrence of the article of interest. If the article was systematically distributed, we would expect to find it in most, if not all the referential contexts that it is expected to occur in. Its absence in such contexts should follow a pattern that can be described as a rule or accounted for by other, additional factors.

Another complication that arises once frequency of occurrence is taken into account as a proxy for the systematic distribution of an article in certain referential contexts is that languages are dynamic and are constantly changing.¹⁰ Thus, there will be many in-between cases of quasi-articles or article candidates such as the one of *ja* in Tz'utujil described above, which may or may not develop into articles or develop from articles into more general nominal markers. It is therefore difficult, if not impossible, to define hard frequency thresholds for what should and should not count as an article, even in cases when corpus data is available and a quantitative analysis would be possible.

In the present study, I included only those markers that appear to be systematically distributed based on the material and evidence available, excluding markers in a conservative way when the evidence was not convincing or conclusive enough. In grammars, for instance, I also considered the distribution of a marker in the relevant referential contexts in examples outside of the chapter discussing articles or noun phrases. In addition, many of the grammars used in this study provide an additional text collection. Whenever such a corpus was available, I only included the article in this study if it also occurred systematically there. As for primary data, I only included markers when the consultants systematically required the presence and rejected the absence of the marker in the relevant referential contexts.

2.3 Summary

In this chapter, I presented the necessary criteria to define articles as a crosslinguistic category, namely their referential function, their occurrence together with nouns, and their systematic distribution. In Section 2.2.2, I showed different referential functions of articles. I also argued that only markers whose main function it is to encode referentiality should be treated as articles. In Section 2.2.3, I argued that articles have to occur together with nouns in the noun phrase, and I

¹⁰There are a number of different types of frequency; for the sake of brevity, I only refer to the token frequency of the article in certain types of referential contexts here.

showed that further requirements with regard to their form or position are not necessary. Section 2.2.4 was concerned with the distributional criterion of articles as a crosslinguistic category. I argued that articles need to be distributed in a systematic way, allowing for exceptions. I showed that we commonly find two types of exceptions: articles can either be absent in contexts in which their presence is expected based on their referential function or they can be present in referentially unexpected contexts. Systematicity can capture this distribution, allowing for exceptions that can be described as rules or strong correlations.

Chapter 3

Referentiality and types of articles

This chapter proposes a definition of each of the 10 article types that we find in the languages of the world. Since they are defined on the basis of their referential functions, Section 3.1 introduces the relevant background concepts related to referentiality. In Section 3.2, I then define the relevant referential functions which serve as the basis for the definition of the 10 article types in Section 3.3. The last part of this chapter, Section 3.4, is a synthesis of the relation between the referential functions and the article types introduced, presented in the light of referential scales, hierarchies, and other structured relations between referential functions from the literature.

3.1 Referentiality

Referentiality is understood as a linguistic category of reference, which concerns the link between a linguistic expression to an abstract object in the discourse, namely the discourse referent. The values of referentiality, i.e. the functions that referring expressions have, are referential functions. In formal semantics, referentiality has usually been approached from the sentence level. The study of referentiality restricted to the sentence level is traditionally located within semantics, also labeled denotational semantics (cf. Kroeger 2018: 17), while discourse reference has been viewed as belonging to pragmatics (e.g. Givón 1981, 1984, Hopper & Martin 1987). In this study, I do not distinguish these two levels of sentence and discourse referentiality and their location within semantics or pragmatics. In doing so, I follow Hawkins, who remarks: “By looking at the total picture, instead of arbitrarily excluding either pragmatics or logical considerations, one is led more easily to the formulation of significant generalisations” (Hawkins 1978: 90).

3.1.1 The discourse situation

This section introduces a schematic discourse situation in order to lay the ground for the distinction between different referential functions, which in turn is the basis for the distinction between different types of articles. The relevant notions are: the discourse situation, the participants (the speaker S and the hearer H), and the discourse referents. To illustrate how the discourse referents relate to the participants and the discourse itself, consider the following example in (68), where speaker S and hearer H are chatting at the bus stop, waiting for the bus to come. A dog approaches, looks around and runs away. S is then saying to H:

(68) The dog seemed lost.

The discourse situation in (68) can be characterized in terms of the context and the discourse participants, i.e. the speaker S and the hearer H. The sentence uttered by the speaker contains the linguistic expression *the dog*, which is linked to the discourse referent R_{dog} . Figure 3.1 shows how such a discourse situation can be represented schematically.

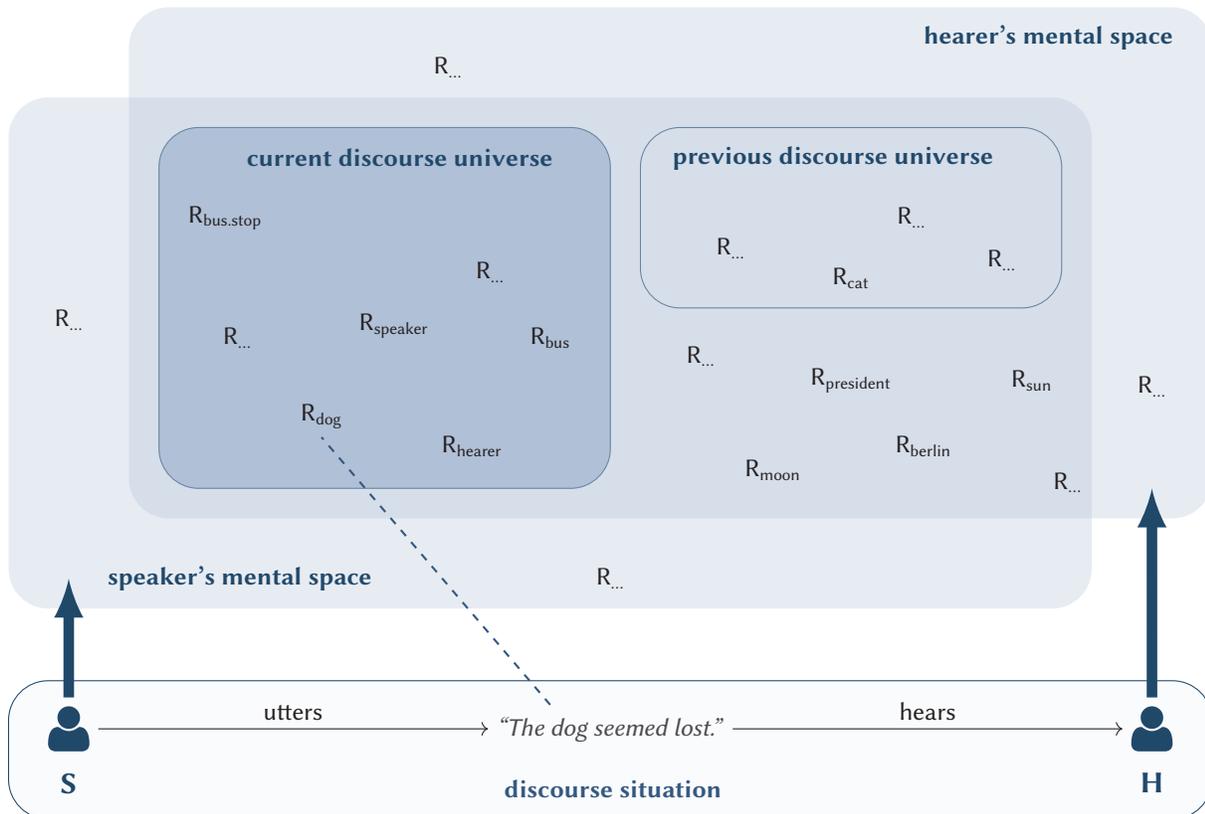


Figure 3.1: Discourse situation and discourse universe

For reasons of simplicity, I take a discourse situation to involve two discourse participants (the speaker S and the hearer H). The linguistic expressions (words) that the speaker utters are also part of the discourse situation; in this case it is the utterance of the sentence *The dog seemed lost*.

In order to properly distinguish between form and function, discourse referents (henceforth: referents) are represented on a different level than the referring linguistic expressions. Referents are defined as (abstract) elements in what I call the speaker's and the hearer's mental spaces. The concept of "mental space" goes back to Fauconnier (1994, 2007) and aims at being a semantic but cognitively plausible concept. Fauconnier (1997: 11) defines mental spaces as "[...] partial structures that proliferate when we think and talk, allowing a fine-grained partitioning of our discourse and knowledge structures." In other words, mental spaces contain our knowledge, beliefs and attitudes of the things we experience and talk about (Langacker 2008: 41). What is important for the present study is that if a referent is part of the mental space of a discourse participant, then the referent is identifiable by that discourse participant. Applied to the discourse situation in (68), the referents R_{dog} and $R_{\text{bus.stop}}$ are part of both the speaker's and the hearer's mental spaces because they are visible or physically perceivable and thus identifiable by both the speaker and the hearer. In other words, those referents are elements in the intersection between the speaker's and the hearer's mental spaces.

The referents, even though being elements in the abstract sets of mental spaces, are linked to the linguistic expression uttered by the speaker. In example (68), the referent R_{dog} is linked to the expression *the dog*. As is schematically illustrated in Figure 3.1, the intersection of their two mental spaces contains the discourse universe which contains all the referents from the discourse situation that both the speaker and the hearer can identify. This includes the referent R_{bus} , even though no bus is necessarily present in the discourse situation. However, the immediate discourse situation given by the bus stop makes the bus that the speaker and the hearer are waiting for (assuming for the sake of simplicity that both are waiting for the same bus) the only situationally salient or relevant bus. This in turn makes the referent R_{bus} part of the current discourse universe. We can also say that the speaker and the hearer evoke the two referents R_{speaker} and R_{hearer} as part of the current discourse universe.

The discourse universe may sometimes be only a subset of the intersection of the speaker's and the hearer's mental spaces because both discourse participants might share knowledge about other referents based on previous shared experience or general and world knowledge. For instance, the discourse participants in (68) may know each other well and thus may have shared many previous discourse situations. This is why Figure 3.1 shows a previous discourse universe within the intersection of the speaker's and the hearer's mental spaces. Of course, there may not only be a single, previous discourse universe but many different ones from different shared experiences. Also, if the speaker and the hearer do not know each other and have never spoken to

each other before, they may have shared referents by virtue of living in the same community; for instance, they may share a referent of a certain park or station in the neighborhood in their mental spaces. Shared knowledge or experience on a larger scale such as being in the same country or belonging to the same religion or culture may also lead to referents being shared between two discourse participants' mental spaces without them having to be acquainted with each other.

Hence, the current discourse universe, related to the current discourse situation, may not contain all the referents that are mutually identifiable by the speaker and the hearer. There can be additional mutually identifiable discourse referents that are neither linked to a physical object present in the discourse situation (i.e. deictic referents) nor previously mentioned within the same discourse situation (i.e. anaphoric referents). In Section 3.2 this will be shown to be important for the definition of certain referent types, namely recognitional as well as absolutely and contextually unique referents. Furthermore, as the two sets of mental spaces only overlap, they allow for each discourse participant to have other referents in their mental spaces that are not identifiable by the other discourse participant(s).

Early views on referentiality and definiteness based on uniqueness (cf. Russell 1905, Strawson 1950) assumed that referentiality was tied to the existence of a referent in the real world. This is problematic in contexts such as the famous sentence *The king of France is bald* in Russell (1905). If no *king of France* exists in the real world, the truth value of the proposition cannot be assessed. Formal approaches aimed at resolving this issue with what is called "possible worlds semantics".¹ In possible worlds semantics, the truth value of a proposition about a referent can be interpreted with respect to any possible world. The approach that I follow in this study does not tie denotation to existence as such and follows a different tradition arguing for referents being interpreted with respect to a discourse universe instead of the real world or some other possible world. To this end, Karttunen (1968, 1976) introduced the notion of the "discourse referent". Thus, referentiality is always interpreted with respect to the discourse itself, for which Givón (1984) used the notion of discourse universe.² Givón (1984: 438) argues that representing referents on the level of the discourse universe does not only make the analysis of referring expressions simpler, but is actually grounded in language use itself. As he puts it:

In human language, it seems, reference–Russell's denotation–is not a mapping from referring linguistic expressions to individuals existing in the RW [real world]. Rather, it is a mapping from linguistic expressions to individuals established verbally in the Universe of Discourse. (Givón 1984: 438)

¹Cf. Carnap (1947), Hintikka (1967), Kaplan (1989), Kripke (1959), Lewis (1970), Montague (1970).

²Here I use Givón's notion of discourse universe. Other related notions and concepts are "common ground", used in Krifka & Musan (e.g. 2012a) and going back to Karttunen (1974), Lewis (1979), Stalnaker (1974). This notion is closely related to the broader concept of "grounding" in Cognitive Grammar (cf. Clark & Brennan 1987, Langacker 1987: 126-128, Langacker 2001, Langacker 2002, Langacker 2008: Chapter 9). Similarly, Hawkins (1978, 1991) makes use of a set called "P-set" (pragmatic set) that contains all referents known to both speaker and hearer.

To illustrate this, Givón (1984) provides the following examples:

- (69) a. *The present king of France is not bald.*
b. *The present queen of England is not bald.*

(Givón 1984: 438)

- (70) a. *I rode a unicorn yesterday.*
b. *I rode a horse yesterday.*

(Givón 1984: 438)

We see in (69) and (70) that even though the sentences have a different interpretation with respect to their truth value in the real world (the referents in (69a) and (70a) do not exist, as opposed to the ones in (69b) and (70b)), the English language does not treat them differently. Both referents in (69) are expressed by the definite article, while both nouns in (70) occur with the indefinite article.

3.1.2 Identifiability, definiteness, and indefiniteness

Traditionally, the literature on definiteness has been split between two major approaches based on uniqueness and familiarity. In this section, I argue that the concept of identifiability can be related to both uniqueness and familiarity and that it can be used to distinguish between different referent types in the definite domain and specific and nonspecific referent types in the indefinite domain.

Identifiability has to do with the knowledge of the speaker and the hearer about certain properties of the referent. Note that the concept of identifiability has received various labels in the literature, e.g. “familiarity” (e.g. Christophersen 1939), “givenness” (e.g. Gundel et al. 1993), “accessibility” (Ariel 1988: e.g.), or “knowledge”. I use “identifiability” in this study, but it could in principle be replaced by any of the other labels.

In Section 3.1.1, identifiability was introduced as the property of discourse referents that are part of the speaker’s and the hearer’s mental spaces. Mutually identifiable referents by both the speaker and the hearer were represented as abstract elements in the intersection between their mental spaces, in the set called the discourse universe. Identifiability itself represents knowledge about the referent and can be understood in the sense of “manifest” as defined in Sperber & Wilson (1986: 39), taken up by Hawkins (1991). A definition of identifiability based on Sperber & Wilson (1986: 39) is presented in (71).

- (71) *Identifiability*

A fact is identifiable by a discourse participant at a given time if and only if she is capable at that time of representing it mentally and accepting its representation as true.

Identifiable facts denote online-knowledge and commitment to that knowledge by a discourse participant. I will use “know” and “knowledge” as well as “identify” and “identifiability” in that sense. Combining the notion of identifiability with the discourse model that was presented in the previous section, we can define identifiable referents and mutually identifiable referents as follows:

(72) *Identifiable referent*

A referent evoked by a referential expression is identifiable by the discourse participant if the referent is an element of her mental space.

(73) *Mutually identifiable referent*

A referent evoked by a referential expression is mutually and unambiguously identifiable by the speaker and the hearer if the referent is an element of both of the speaker’s mental space and the hearer’s mental space, and if the speaker and the hearer assign the same referent to the referring expression.

As various authors have pointed out, the discourse universe is not a stable and fixed structure but dynamic and constantly updated during the discourse (e.g. Giannakidou 1998, Hawkins 1991, Krifka & Musan 2012b, Langacker 2008). Hawkins (1991) shows that, even though a referential expression is linked to a referent that cannot be fully identified by the hearer based on her knowledge at the point of utterance, the hearer is trying to establish identifiability if the referent is presented as definite. This capacity helps to construct and update the discourse universe and has been labeled “accommodation” (Karttunen 1974, Stalnaker 1974, Lewis 1979, Krifka & Musan 2012b). It corresponds to a repair strategy to rescue the interpretation of an utterance. Hawkins (1991: 413) illustrates hearer accommodation using the following example, uttered in a context in which Mary and Peter are looking at a landscape where she has noticed a distant church.

- (74) (Context: Mary and Peter are looking at a landscape where she has noticed a distant church. She says to him:)
*I’ve been inside **that** church.*
(Hawkins 1991: 413)

Hawkins (1991: 411) points out that the use of the demonstrative in such an utterance is fully acceptable:

The speaker is simply telling the hearer to extend the relevant P-sets [here: discourse universe] on these occasions, and expects the hearer to accept these extensions, presuming that they are at least compatible with what he knows and that he has no reason not to accept them.
(Hawkins 1991: 411)

It is not required for Mary to ensure that Peter first sees the object ‘church’ in the real world to be able to identify the referent R_{church} linked to the expression *that church*. Due to the marking of the referent R_{church} as definite and thus identifiable, Peter tries to establish identifiability relying on the fact that Mary marked the referent as definite because he can identify it as well. The fact that hearer accommodation is possible has two consequences concerning the nature of the discourse universe: on the one hand, it needs to be dynamic and updatable. On the other hand, a referential marker does not only mark identifiability but it can also establish a referent as identifiable.

Another property of the concept of identifiability is that it is context-sensitive, which means that depending on the context, the speaker’s and the hearer’s knowledge of a referent can be sufficient for full identification or not. It seems to be commonly assumed that spatial deixis (or physical perception in general) is sufficient for a referent to be identifiable.³ In this vein, nouns accompanied by demonstratives are taken to be definite, since demonstratives typically involve deixis, which entails knowledge of physical properties based on physical perception such as visibility. For the majority of instances, it holds true that visibility licenses mutual identifiability by the hearer and the speaker. There are, however, counter-examples, illustrating that identifiability must be context-sensitive and is not automatically licensed by visibility. In example (75), the referent R_{boy} of the expression *a boy* is linked to the visible object of ‘a boy’ in the discourse situation. The discourse context of (75) is the following: S and H are on a ship in the ocean, looking at the water, when S spots a boy in the water. S then says:

(75) *Look! There is a boy in the water!*

As the expression referring to R_{boy} includes the indefinite article *a*, the referent R_{boy} is marked as non-identifiable, which in turn suggests that visibility is not always sufficient to license full identifiability.

The main focus in semantics with respect to referentiality has probably been on definiteness, mainly divided into uniqueness and familiarity approaches. Uniqueness-based accounts of definiteness go back to Russell (1905) and basically require a definite description to be unique. In Russell (1905), this uniqueness condition is directly built into the truth condition of a definite description. The meaning of the famous sentence *The King of France is bald* equals the conjunction of the following three statements: (i) there is a King of France and (ii) there is only one single King of France and (iii) the King of France is bald. The quantificational aspect of the definite article (the fact that uniqueness is part of its denotation) has been subject to criticism, so that a more “moderate” version of the denotation was proposed in Frege (1892) and Strawson (1950). They integrate uniqueness as a presupposition for the felicitous use of the definite description (as

³Note that I treat referents and objects in the real world as different elements on different levels. Thus, strictly speaking, a referent in this sense can never be perceived directly, it can only be linked to an object from the world that is perceivable (cf. Section 3.2).

a precondition for the element that must be met for its successful use) instead of including it in its denotation. With regard to the example about the King of France, this means that there has to be one and only one King of France, and if the King of France is bald, the proposition is true. If he is not bald, the proposition is false. If there is no unique King of France, the truth value of the sentence is not defined.

Independently of how exactly uniqueness is used to define definiteness, it straightforwardly accounts for those cases in which the discourse referent marked as definite somehow stands out as a single prominent referent. However, two interrelated problems arise with this account. On the one hand, uniqueness as such does not provide any information on the conditions under which a certain referent can be unique. On the other hand, referents can be marked as definite even though they are not unique in the entire world, which calls for some mechanism to restrict uniqueness to a certain relevant domain. Heim & Kratzer (1998) resolve the second problem by requiring that a referent expressed by a noun and a definite article be unique in a contextually salient subset of the world. However, this still does not make explicit which conditions have to hold for the referent to be salient. Therefore, also this version of contextual uniqueness needs an additional mechanism that regulates it with respect to the discourse situation.

Anaphoric and therefore salient but not necessarily unique referents have received a lot of attention in the semantics literature, which has led to a competing approach to definiteness, namely the familiarity approach going back to Christophersen (1939). He points out that instead of uniqueness, rather familiarity with (or non-ambiguity of) the referent should be the prerequisite for definiteness. This was taken up by Karttunen (1968) and the influential model of File Change semantics (a discourse model that accounts for definite and indefinite uses of referents) in Heim (1988). What exactly makes a referent familiar differs from account to account; in Kamp (1984) and Heim (1988, 2002), familiarity is licensed by previous mentioning in the discourse, i.e. anaphoric uses. This is also taken to be the distinctive property of definites as opposed to indefinites:

A definite is used to refer to something that is already familiar at the current stage of the conversation. An indefinite is used to introduce a new referent. (Heim 2002: 223)

Without going into details of the formalization of this framework here, it should be mentioned that it offers a precise, cognitively grounded mechanism to account for anaphoric definites and newly introduced indefinite referents, which uniqueness-based approaches could not deal with in a straightforward way. Yet there are contexts with definite referents that this framework cannot deal with either. Consider the examples in (76):

- (76) a. *The man drove past in a car. **The horn** was blaring loudly.*
b. ***The professor** that we were just talking about ...*

- c. *Pass me **the bucket** that is over there.*
(Hawkins 1991: 410)

Although the discourse referents in (76) are newly introduced in the discourse, they are expressed by a noun together with the definite article, marking the referents of these nouns as definite. This shows that also the purely familiarity-based approaches such as File Change semantics cannot capture the variety of contexts that involve definiteness, as they focus on the (dynamic) semantic status of the discourse referent, but not on the discourse participants, their relation to the referent, their intentions etc.⁴

Therefore, I argue that uniqueness or familiarity alone are not sufficient to account for all types of definite contexts and to explain the distribution of definite expressions (and thus the distribution of different types of articles) in the discourse. The concept of familiarity seems to be rather restrictive in that it focuses on anaphoric definites. Thus, one would need to add an entirely different mechanism to capture the different non-anaphoric contexts of definites. Uniqueness, on the other hand, is less restrictive, but not applicable in a straightforward way to natural language use, since the domain of uniqueness as restrictive and relativizing factor must be implemented in some way to make it compatible with different discourse situations. However, a slightly different interpretation of the notion of “uniqueness” was proposed that is more apt to be applied to various types of definites in natural language use. As was already noted by Christophersen (1939) and emphasized in Löbner (1998), uniqueness can be understood in terms of unambiguous identifiability:

Though the previously acquired knowledge may relate to the very individual meant, yet it is often only indirectly that one is familiar with what is denoted by the word. It may be something else that one is familiar with, but between this ‘something’ and the thing denoted there must be an unambiguous relation. (Christophersen 1939: 72f)

Here it is: the crucial ‘unambiguous’, i.e. 1-to-1, relationship between the trigger and the associate, based on general knowledge. (Löbner 1998: 12).

This is exactly how I use the notion of identifiability in this study, i.e. as unambiguous identifiability. As mentioned above, if a referent is mutually and unambiguously identifiable, both the speaker and the hearer are capable of representing it mentally and of accepting its representation

⁴Roberts (2003) elaborates on the familiarity requirements proposed in Heim (1988), distinguishing between two levels of familiarity. Without going into detail, weak familiarity subsumes discourse referents that are contextually accessible or given by world and cultural knowledge, while strong familiarity refers to anaphoric uses. This again shows that in order to make the concept of familiarity applicable to more contexts involving definites, a more elaborate mechanism has to be applied. Although being able to capture phenomena such as bridging, this version can still not account for those uses of definites that uniqueness can do.

as true, and both assign the same referent to the referring expression. A referent can be unambiguously identifiable because it is a familiar referent such as an anaphoric referent. On the other hand, an “unfamiliar” referent can nevertheless be unambiguously identifiable if it is the only salient referent of its kind in a given context, either because it is unique or because it is contextually unique, for instance, by an unambiguous relation to another familiar referent, which we see in case of bridging referents. Therefore, uniqueness in the sense of an unambiguous relation and identifiability are equivalent concepts which I use in Section 3.2 to define different referential functions in the definite domain. This also means that uniqueness and familiarity both play a role in the definition of unambiguous identifiability as used here, while neither of the two are sufficient notions to define all referent types in the definite domain.⁵

In other words, referent types in the definite domain need to be mutually and unambiguously identifiable by the speaker and the hearer, i.e. they are elements of the intersection of the speaker’s and the hearer’s mental spaces. It is their unambiguous identifiability that guarantees that they are the only salient or relevant referent of their kind.⁶ What does it mean that a referent is unambiguously identifiable by a discourse participant? I will relate unambiguous identifiability to two different but related concepts used in pragmatics and semantics: the maxim of Quantity (Grice 1975, Levinson 1987) and the maximality condition for definite plurals (Link 1998).

Going back to the maxim of Quantity (Grice 1975: 45-46), the Q-principle (Levinson 1987: 401, Hawkins 1991) ensures that the speaker does not say less than she knows, and that the hearer in turn can rely on that:

(77) *Q-principle:*

Speaker’s maxim: Do not provide a statement that is informationally weaker than your (or the hearer’s) knowledge of the world allows.

Hearer’s corollary: Take it that the speaker made the strongest statement consistent with what the speaker and the hearer know.

(based on Levinson (1987: 401))

This is relevant to unambiguous identification since it means that the speaker refers to the exact number of referents that are mutually identifiable, neither more nor less. The hearer, on the other hand, can rely on the fact that when the speaker utters *the dog seemed lost*, then there is at least one and no more than one identifiable referent of the kind ‘dog’.

In a way, the Q-principle is an informal and discourse-oriented version of the maximality condition in formal semantic approaches used for definite plurals (e.g. Schwarz 2012, Abbott

⁵The fact that familiarity, uniqueness, and identifiability (in the traditional sense) are not sufficient on their own but all necessary to define different types of definite referents is also noted and discussed in more detail by Lyons (1999: 2-15).

⁶In the remainder of this study, I use “identifiable” and “identifiability” in the sense of mutual and unambiguous identifiability.

2010: 159-160).⁷ Such a condition is necessary as the uniqueness condition requires for a definite expression that there be only a single salient referent, which can no longer hold as such for a group of referents. Thus, instead of referring to a single contextually salient referent, plurals refer to the maximality of contextually salient referents of a given kind. An utterance like *the dogs seemed lost* thus refers to all referents of the kind ‘dog’ that are contextually salient.

Davis et al. (2014) illustrate the effects of the Q-principle or the maximality condition using the following examples:

- (78) a. Context: one bear
The bear attacked.
- b. Context: more than one bear
The bears attacked.
(Davis et al. 2014: 200)
- (79) Context: two identical cups sitting next to each other
- a. # *Pass me the cup.*
- b. *Pass me the cups.*
(Davis et al. 2014: 200)

Example (78) shows the successful application of definite singular and plural expressions. If there is only a single contextually salient referent as in (78a), then the singular is used to refer to the maximal number of contextually salient referents of the kind ‘bear’, i.e. a single one. In (78b), on the other hand, the use of the plural definite *the bears* to refer to more than one contextually salient referent of the kind ‘bear’ has to be interpreted as referring to all contextually salient referents of the kind ‘bear’. Example (79) on the other hand shows how the singular definite expression *the cup* is not felicitous in a context featuring more than a single salient referent of the kind ‘cup’.

In that sense, uniqueness applied to singular referents is only a special case of maximality; it refers to all referents of a given kind that are contextually salient, only that “all referents” amount to a single one. Link (1998: 184) explains this in the following way:

Now in the plural domain two concepts for expressing the notion of exactness have to be distinguished that coincide in the singular case: uniqueness and maximality. While we have to give up the former, we can still postulate the latter [...]. Thus, exactness does not mean that there is only a single individual in the domain with the property in question, but rather that there is a maximal one among them, and this unique element is picked out.

⁷Another variant of this condition is the concept of “Inclusiveness”, proposed by Hawkins (1978: 157.-167).

This means that identifiability needs to be understood in terms of maximality. Therefore, a mutually identifiable referent corresponds to the maximal number of referents of a given kind which is mutually and unambiguously identifiable by the speaker and the hearer.

The next step is to make explicit how referents can become mutually and unambiguously identifiable by the speaker and the hearer. To that end, I will define 8 referential functions on the basis of how their mutual and unambiguous identifiability is achieved in Section 3.2. These are all functions in the definite domain. Note that the definite domain or definiteness should not be understood as a primitive referential type or function. This is very important for the definition of articles, given that certain types of articles can only express a subset of the functions in the definite domain.

Indefiniteness, even more than definiteness, is usually discussed in combination with a specific linguistic form, the indefinite article *a* in English in most cases. Heim (1991: 514-515), for instance, argues that the indefinite article, in addition to the existence presupposition which it shares with the definite article, has a non-uniqueness condition. She bases that on examples such as (80), for which she argues that the indefinite article is not felicitous because it is known that there is can only be single referent.

- (80) a. *The (#a) head of the victim had been shaved.*
b. *The (#a) weight of our tent is under 2kg.*
(Heim 1991: 514)

Hence, to make the conditions on the indefinite article more restrictive and distinct from the one of the definite article, Heim (1991: 32) argues that the indefinite article cannot be used with a referent for which uniqueness is already presupposed, as for *weight* and *father* in (80) above. Note that the examples in (80) can also be accounted for by a more elaborate distinction of semantic noun types or concepts as proposed in Löbner (1985, 2011). In this approach, nouns such as *father* and *weight* are classified as functional nouns, which means that they have an argument that inherently relates them to a different referent and makes them contextually unique and thus definite, explaining the incompatibility with the indefinite article. Therefore, Löbner accounts for the infelicitous use of the indefinite articles in examples like (80) by having uniqueness inherently built into the semantic properties of the referents.

Nonetheless, other scholars have argued against uniqueness being relevant for the use of the indefinite article. Hawkins (1991: 420-422) shows that the indefinite article can implicate uniqueness in some contexts, whereas in others, it cannot. Examples of such contexts are shown in (81):

- (81) a. *There is **a pizza** in the fridge, and **a cake** in the pantry.*

- b. *Pass me a bucket. ~Pass me one of the buckets (before us).*
(Hawkins 1991: 421, 417)

In (81a), the referents of *pizza* and *cake* are marked as non-identifiable by the indefinite article. The latter is required by the existential construction and the referents are understood as contextually unique. Example (81b), on the other hand, provides evidence for non-identifiable referents marked by the indefinite article to be contextually non-unique. Therefore, Hawkins concludes that the meaning of the indefinite article is neutral to uniqueness. In some contexts, this corresponds to the ambiguity of the indefinite expression with respect to uniqueness. In (82), the referent of *a movie* is not necessarily contextually unique (Mary could have watched another movie on that night), but can be interpreted as such (Mary did not see another movie that night).

- (82) *A movie that Mary was watching last night was really interesting.*
(Hawkins 1991: 422)

I follow Hawkins in that I do not consider (non-)uniqueness as an important condition for the use of the indefinite article. However neither Heim's nor Hawkins's approach accounts for indefiniteness independently from linguistic expressions. Instead, both aim at explaining the meaning(s) of the indefinite article in English, which, again, is not sufficient in order to define and compare articles across languages. In the following section, I argue that indefiniteness, like definiteness, should not be treated as a primitive referent type either. It rather subsumes the two types of specific and nonspecific referents.

To sum up this section, we can use mutual and unambiguous identifiability to distinguish between the definite and the indefinite domain. The mutual and unambiguous identifiability of a referent can be achieved through the familiarity of a referent, as is the case with anaphoric referents. However, the condition that the referent needs to be unambiguous identifiable relates to the uniqueness requirement of definite referents. Unambiguous identification means that in a given discourse situation, there is only a single salient referent of its kind evoked by the referential expression. If reference is made to a set of more than one identifiable referent of its kind, uniqueness can be translated into the maximality of contextually salient referents of their kind in a given discourse situation. Lastly, we saw that unambiguous identifiability has to be mutual between the speaker and the hearer, meaning that both have to be able to unambiguously identify the same (set of) referent(s), assigning the same (set of) referent(s) to the referential expression used.

3.2 The referential functions

This section defines the referential functions that are necessary to distinguish between different article types. We can divide all the relevant functions into the definite and the indefinite domain. In the definite domain, the following functions need to be distinguished: deictic, anaphoric, recognitional, establishing, situationally unique, contextually unique, bridging, and absolutely unique. The indefinite domain consists of specific and nonspecific functions. The following sections will introduce these in turn.

3.2.1 Deictic

A spatial deictic referent is identifiable because it is linked to an object in the real world that is present in the discourse situation. Henceforth, I will use “deictic” only in the sense of “spatial deictic”. The most prototypical expression marker of a deictic referents is a spatial demonstrative, as is shown in (83):

(83) *Look at **that beautiful house!***

The presence of the object in the discourse situation makes the referent the single most salient referent of its kind in the discourse universe. Therefore, the referent is unambiguously identifiable by both the speaker and the hearer. A deictic referent can thus be defined as follows:

(84) *Deictic referent* R_{dei}

A referent evoked by a referential expression is a deictic referent R_{dei} if it is unambiguously linked to an object in the discourse situation so that both the speaker and the hearer can unambiguously identify the referent and assign the same referent to the referential expression.

When the speaker utters (83), we can imagine that both the speaker and the hearer can already see (or rather physically perceive) the object that the expression refers to. However, Hawkins (1978: 113) notes that the object linked to the referent does not have to be visible to the hearer (yet). In such cases, using the demonstrative to express a deictic referent, the speaker rather instructs the hearer to try to see or perceive the object linked to the referent in question. Moreover, Hawkins (1978: 112-113) constructs a scenario in which the speaker cannot perceive the object linked to the referent directly but only has indirect evidence, whereas the hearer can see the object. The context of this scenario is given in (85a), and the utterance containing a deictic, speaker-invisible and hearer-visible reference is shown in (85b). Note that a deictic referent, requesting the hearer to physically perceive the object linked to the referent, does not necessarily have to be expressed by a demonstrative. In (85b), it is the definite article which codes the referent as a deictic one.

(85) a. Context: Suddenly, a man wearing prisoners' uniform jumps down from the wall and starts running for his life. The policeman hears a frantic voice from the other side saying:

b. *PC 49, catch **the jailbird!***

The fact that deictic referents involve physical perception and that they are not necessarily expressed by demonstratives is also shown in example (86), repeated from Section 3.1.2:

(86) *Look! There is **a boy** in the water!*

In Section 3.1.2, it was argued that physical perception is not always a sufficient condition for unambiguous identifiability, which accounts for the use of the indefinite article with the deictic referent in example (86). For the purposes of this section, it is important to note that also the indefinite article can code a deictic referent, as long as it is assigned to an object in the immediate discourse situation that either can be or should be physically perceived by the hearer.

3.2.2 Anaphoric

A very important and widely discussed subtype of definite contexts are anaphoric definites. They have been central in works on definiteness and discourse reference from various theoretical viewpoints. The formal mechanism of File Change semantics (Heim 1988), for instance, restricts the notion of definiteness to anaphoric uses. Another approach to definiteness and identifiability in which anaphoric uses play a major role is found in Ariel (1988, 1990, 2001). She examines the correlation between the linguistic expression of the discourse referent (NP, pronoun, demonstrative, etc.) and the distance between the antecedent and the anaphoric referent. This is also a relevant factor for the distribution of definite articles. Since articles generally occur together with nouns and not with pronouns, the variation between pronominal and nominal expressions of referents also restricts the distribution of definite articles in anaphoric contexts. In this vein, Epstein (2002: 340) pointed out that neither uniqueness nor familiarity can account for the distribution of the definite article with respect to other definite descriptions or for its infelicitous use in examples such as *There's a cat in the yard ...It (#The cat) is eating a mouse*. The Accessibility Hierarchy in Ariel (1988) aims at explaining such cases. Based on empirical findings, the definite article could be classified as low accessibility marker, which means that it is not felicitous in contexts in which the referent is topical (or highly accessible). Since it is not possible to formulate general deterministic rules for the conditions that trigger the use of definite articles vs. pronominal forms, I will not discuss this here. I define anaphoric referents R_{ana} in the following way:

(87) *Anaphoric referent R_{ana}*

A referent evoked by a referential expression is an anaphoric referent R_{ana} if it is identi-

cal to another previously mentioned referent R_{ante} , which is part of the current discourse universe, so that both the speaker and the hearer can unambiguously identify the referent R_{ana} and assign the same referent to the referential expression.

This means that the anaphoric referent R_{ana} requires to be identical to another referent R_{ante} , the antecedent, which is part of the discourse universe as it was uttered at an earlier point in time in the same discourse situation. From their shared identity, it follows that the anaphoric R_{ana} is an element of the discourse universe as well, which makes it mutually and unambiguously identifiable by both the speaker and the hearer.

3.2.3 Recognitional

Recognitional referents represent a type of definite referents that are elements of the intersection between the speaker's and the hearer's mental spaces without being an element of the current discourse universe. Recognitional referents are identifiable by both the speaker and the hearer because of previous shared experience or specific knowledge. Such uses have been discussed in the previous literature mainly as one subtype of other referential functions (Auer 1981, 1984, R. Chen 1990, Gundel et al. 1993, Diessel 1999: 105). The probably most detailed account of recognitional referents is given in Himmelmann (1996: 230-239) and Himmelmann (1997: 61-82).⁸ An example of a noun with a recognitional referent can be seen in (88).

- (88) [Context: S and H, friends, are talking about the past when they used to live in the same flat. Suddenly, S remembers that they used to have a cat.] *What happened to **the cat** (we used to have)?*

The referent assigned to the expression *the cat* is identifiable and marked as such without having been mentioned previously in the discourse and without being linked to any object that would be physically present or perceivable in the current discourse situation. Instead, they are instead marked as identifiable because they are based on personal or specific knowledge and experiences shared by the speaker and the hearer (Himmelmann 1996: 230). Here, specific knowledge should be understood as opposed to general (world) knowledge relevant to unambiguously identify contextually unique referents. While recognitional uses do not track referents through the discourse like anaphoric use do, both referential functions nevertheless share certain properties in that the referents may have been mentioned in other previous discourse situations. In such cases, we can represent recognitional referents as being part of a previous discourse universe, as was shown in Figure 3.1 in Section 3.1 for the referent R_{cat} .

⁸In Himmelmann (1997), recognitional referents are discussed under the label of “anamnestic” uses.

As can be seen from examples (88) and (89) below, recognitional referents often involve additional descriptions of the referent, e.g. in the form of relative clauses or other modifying elements of the noun expressing the referent.

- (89) a. *I couldn't sleep last night. **That dog** (next door) kept me awake.*
(Gundel et al. 1993: 278)
- b. *...it was filmed in California, **those dusty kind of hills** that they have out here in Stockton ...*
(Himmelman 1996: 230)

Himmelman (1996) offers the following explanation for the frequent use of clausal modifiers such as relative clauses with recognitional referents, adapting the account of Auer (1981, 1984) for the recognitional uses of the German demonstrative *dies* 'this'.

"A central feature of this use is that the speaker anticipates problems with respect to the information used in referring to a given referent. That is, the speaker is uncertain whether or not the kind of information he or she is giving is shared by the hearer or whether or not this information will be sufficient in allowing the hearer to identify the intended referent."
(Himmelman 1996: 230)

In addition, Himmelman (1997: 58) notes that the expression of recognitional referents are often followed by pauses and reassurance markers, allowing the hearer to request more information from the speaker if she is not able to unambiguously identify the referent. Another typical context of recognitional uses, causing the speaker to doubt the successful identification of the referent by the hearer, is given when the speaker fails to come up with an appropriate referring expression that would allow for identification of the referent by the hearer (Himmelman 1996: 234). The speaker may then use a lexically more general expression together with the recognitional marker (e.g. a demonstrative in English) to fill in until the speaker is able to repair the expression or until the hearer asks for elaboration. An example of such scenarios is shown in (90), taken from the Pear Stories, where the speaker attempts to refer to the paddle ball that the boy is playing with in the video.

- (90) *and this one's ...playing with one of **those ...those wooden things** that you hit with a ball.*
(Himmelman 1996: 234)

Recognitional referents can thus be defined as follows:

- (91) *Recognitional referent* R_{recog}

A referent evoked by a referential expression is a recognitional referent R_{recog} if it is part

of both the speaker's and the hearer's mental spaces without being part of the current discourse universe. The expression that evokes the recognitional referent may need to include further explanatory information so that both the speaker and the hearer can unambiguously identify the referent and assign the same referent to the referential expression.

Note that dedicated recognitional markers are often used to express mythological referents in stories. Also [Himmelmann \(1997: 68\)](#) observes that in Wubuy and other Australian languages, protagonists of mythical stories are often mentioned explicitly fairly late in the stories. The audience is rather reminded of those protagonists, e.g. by the use of recognitional markers. The data presented in Section 5.4 on recognitional articles confirms this use in different languages spoken in Australia and Papunesia. It is unclear at this point why recognitional marking is often restricted to specific, personal knowledge, and at the same time commonly applied to protagonists in mythological narratives, which would rather count as cultural and more general knowledge within the speech community. One possible explanation is that the use of recognitional markers in the latter type of contexts is an extended use, creating a higher degree of proximity between the narrator (or the speaker) and the audience (or the hearer). This secondary function of recognitional uses has also been suggested outside of the context of Australia or Papunesia in [Diessel \(1999: 106\)](#), who notes: "The recognitional use often implies that speaker and hearer share the same view or that they sympathize with one another." He provides the following examples, adapted from [Lakoff \(1974\)](#), to make this point:

- (92) a. *How's **that** throat?*
b. ***That Henry Kissinger** sure knows his way around in Hollywood.*
([Diessel 1999: 107](#))

In (92a), the use of the demonstrative *that* conveys that the speaker is sympathizing with the hearer in being concerned about the hearer's health. The speaker could be a friend but also the doctor of the hearer, expressing care and creating more proximity between them. By using *that* in (92b), the speaker suggests that the hearer shares her views about Henry Kissinger, which again serves to create a higher level of proximity between the speaker and the hearer in the current discourse situation.⁹

To conclude, a recognitional referent is part of both the hearer's and the speaker's mental spaces without belonging to the current discourse universe. This distinguishes it from a deictic, situationally unique, or an anaphoric referent, which are part of the current discourse universe.

⁹Such uses are called "emotional deixis" in [Lakoff \(1974\)](#) because of the proximity and sympathy they create between the speaker and the hearer.

3.2.4 Establishing

In contrast to the previously discussed referential functions, the identifiability of establishing referents is—as the name suggests—established when the referential expression is uttered.¹⁰ Establishing referents correspond to cases in which a referent is marked as definite, although it is not unambiguously identifiable by the hearer yet, as is shown in example (93). Here, the referent of *the rumour* is established as identifiable by the information provided in the complement clause that follows the head nouns. The latter is marked by a definite article by the speaker, signaling to the hearer that she can or should construct the referent as (soon to be) identifiable.

- (93) *London has been buzzing with **the rumour** that the Prime Minister is going to resign.*
(Hawkins 1978: 102)

From a cognitive perspective Epstein (2002) argues that by using a linguistic expression that marks the referent as definite, the hearer can establish or construct it as such, relying on the fact that she will be able to identify the referent shortly after the utterance.

The pragmatic requirements that allow such a commitment from both the speaker and the hearer are formulated in Hawkins (1991). They are based on the principles of Quantity and Informativeness, which allow the speaker to mark something as definite (i.e. identifiable by the hearer as well) because the speaker can rely on the hearer's commitment to accept the referent as identifiable and to construct identifiability. The hearer, in turn, can rely on the speaker to mark a referent as definite only if it is unambiguously identifiable for the hearer as well or will be shortly after the time of utterance. In other words, the hearer is able to accommodate the referent as identifiable. Establishing referents can be thus defined as follows:

- (94) *Establishing referent* R_{est}

A referent evoked by a referential expression is an establishing referent R_{est} if it is an element of the speaker's mental space, if the speaker's intention is to establish R_{est} as a salient and important referent in the discourse universe, so that R_{est} will be an element of the hearer's mental space and R_{est} will be mutually and unambiguously identifiable and assigned to the same referential expression by the speaker and the hearer.

In addition to constructions using a the complement clause to provide the information required to establish the referent as identifiable, establishing referents are often expressed by a head noun with a following relative clause. Two such examples are shown in (95).

- (95) a. ***The bloke** Ann went out with last night phoned a minute ago.*
(Lyons 1999: 3)

¹⁰This term is taken from Hawkins (1978), who uses it only for establishing relative clauses, i.e. in a more restrictive way than in the present study.

- b. *Mary's gone for a spin in **the car** she just bought.*
(Lyons 1999: 8)

It is important to note that the establishing referents are not defined on the basis of certain syntactic structures such as relative or complement clauses. Example (96) shows how the information establishing the referent as identifiable to the hearer can be expressed in a separate sentence with no syntactic dependency between the noun expressing the referent and the establishing information.

- (96) *Have you heard **the news**? A cabinet minister has just resigned; I didn't catch which one.*
(Lyons 1999: 261)

Thus, establishing referents are not expressed by a definite expression because their referential status as mutually and unambiguously identifiable licenses the use of such an expression. Rather, establishing referents R_{est} are expressed as identifiable by the speaker with the intention to signal to the hearer that she should construct identifiability.

Establishing functions are often expressed by similar syntactic structures as recognitional functions, e.g. nouns with a determiner and a following relative clause containing more information about the referent. Firstly, the two uses differ in the functions they serve; recognitional referents are used to signal that the referent is mutually identifiable based on shared experience or knowledge between the speaker and the hearer. In case the speaker is not able to come up with the appropriate expression for the referent, a recognitional referent also signals to the hearer that she should give feedback to the speaker whether or not the expression used is sufficient for mutual unambiguous reference, or that she should help the speaker to find the appropriate referential expression. Establishing referents, on the other hand, are not (yet) unambiguously identifiable by the hearer; in that case, the speaker rather requests the hearer to construct it as such. Moreover, as argued by [Himmelman \(1996, 1997\)](#), establishing uses do not include pauses or repairs that are typical for the expression of recognitional referents. Thus, even though the expressions of recognitional referents and establishing referents may be similar syntactically, there are clear prosodic differences that come with the expression of these two types of referential functions.

3.2.5 Situationally unique

Situationally unique referents are very similar to deictic referents. [Hawkins \(1978: 111-112\)](#) describes a situationally unique referent as follows: "It is similar to the visible situation use in that the referent exists in the immediate situation in which the propositional act of reference is taking place". He provides the examples shown in (97).

- (97) a. *Don't go in there, chum. **The dog** will bite you.*

- b. *Beware of the dog.*
(Hawkins 1978: 112)

It is the immediate context of the discourse situation that allows for the identification of a situationally unique referent. However, in contrast to deictic referents who require the physical perception of the object linked to them, expressions of situationally unique referents such as *the dog* in (97) only require the hearer to anchor or locate the referent in the immediate discourse situation. In neither (97a) nor (97b) does the referent of *the dog* have to be visible or otherwise physically perceivable. In such examples, the expression of a situationally unique referent rather corresponds to the information that such an object linked to the referent exists in the immediate discourse situation, i.e. ‘a dog’ in the case of example (97). In addition to the information about the existence of an object linked to the referent and the referential expression, the hearer is also “being instructed to ‘locate’ the referent in the immediate situation of utterance. In contrast to the demonstratives he is not being instructed to actually perceive it, but only to assign it to the situation which he is in, in the sense that he understands that it exists in this situation” (Hawkins 1978: 114). Situationally unique referents can thus be defined as follows:

- (98) *Situationally unique referent* $R_{\text{sit.u}}$

A referent evoked by a referential expression is a situationally unique referent $R_{\text{sit.u}}$ if it is unique in the immediate discourse situation so that both the speaker and the hearer can unambiguously identify the referent and assign the same referent to the referential expression.

Hawkins (1978) groups together situationally unique and deictic referents under the cover term of “immediate situation uses”. He groups them together because for both types of referents, the identification of the referent relies on the immediate discourse situation. The only difference is that deictic referents involve some sort of physical perception of an object linked to the referent in the discourse situation, while situationally unique referents are unambiguously identifiable because they are the only referent of their kind in the immediate discourse situation. Hawkins (1978: 111-115), discussing those two types of referents, shows how the use of demonstratives vs. the definite article in English patterns with deictic and situationally unique referents, respectively.¹¹ For the purposes of the present study, I will keep these two referent types separately, as the use of different types of articles in the definite domain suggests that situationally unique functions rather pattern with other semantic definite uses of articles (cf. Chapter 5).

¹¹To be precise, the demonstrative is argued to be restricted to deictic uses in which visual or physical perception is required. The use of the definite article is argued in Hawkins (1978: 111-115) to usually not involve physical perception and express what I call here a situationally unique referent. However, Hawkins (1978: 115) notes that the definite article can under certain circumstances also be used to express a deictic referent: “But whether visibility is simultaneously required by the definite article depends on factors which are extraneous to the meaning of the itself. This important point has not been appreciated in the literature.”

3.2.6 Contextually unique

What I call contextually unique goes back to “larger situation uses” in Hawkins (1978: 115-130). Contextually unique referents are mutually and unambiguously identifiable because they are constructed as unique in a larger context based on (general) knowledge shared by the speaker and the hearer. The discourse referent does not necessarily have to be familiar or visible to the discourse participants. It rather is or is constructed as the only salient referent of its kind in a given context.

A contextually unique referent may be linked to a larger context, e.g. the village, city, or the country that the discourse situation is located in. Two such examples of contextually unique referent linked to discourse contexts which are larger than the immediate discourse situation are given in (99) and (100). Note that in these cases, the speaker and the hearer do not have to know each other. As long as they are currently in the same country, one of two strangers could start a conversation with the utterance in (99). The discourse situation and shared referents would then extend to referents linked to the country that the speaker and the hearer are currently situated in. This can make the referent of *the prime minister* mutually and unambiguously identifiable to both discourse participants.

- (99) *I hear **the prime minister** behaved outrageously again today.*
(Lyons 1999: 3)

Hawkins (1978: 117) notes that there can be two different types of identification that are involved in identifying the referent of *the prime minister* in (99). Identification can either be based on specific knowledge or general (world) knowledge. Imagine that the speaker of (99) is an active member of the opposition party; it is likely that she has specific knowledge about the individual currently serving as the prime minister, knowing the prime minister’s name, her party membership, details of her political opinions, etc. In this case, the speaker has specific knowledge about the individual linked to *the prime minister*, and can thus represent its referent including this knowledge. Now imagine that the hearer is a visitor in the country of the discourse situation, without any specific knowledge about the current political situation. Therefore, the speaker is likely to represent the referent of the *the prime minister* only with very general knowledge about the general function of prime ministers in various countries in the world. Even in such a scenario, the utterance in (99) is felicitous, meaning that the hearer must nevertheless be able to represent and unambiguously identify the referent of *the prime minister*.

The approach of Löbner (1985, 2011) inherently accounts for that. According to Löbner, ‘prime minister’ belongs to the group of functional concepts, which include a possessor argument in their lexical properties. In other words, ‘prime minister’ is inherently linked to another concept such as ‘state’ in an unambiguous way because a state, by default, only has one prime minister at a time. Therefore, the referent of *the prime minister* can be constructed as unique and thus

unambiguously identifiable in the larger context of the country that the discourse is situated in. For such contextually unique uses, Hawkins (1978: 117-122) argues that although discourse participants may have more specific knowledge in addition to general (world) knowledge like the speaker in the scenario of (99) described above, it is the general knowledge that is necessary and sufficient to construct the referent as contextually unique.

Another example of contextually unique referents that are unambiguously identifiable in a larger context built from general world or cultural knowledge is (100).

- (100) (Context: S is visiting a village and is looking for the church. S sees H and asks for help.)
*Where is **the church**?*

Here, the relevant larger context corresponds to the village that the conversation takes place in. Again, the speaker and the hearer do not have to share anything else than general world knowledge about villages in their cultural community in the widest sense. This allows them to assume that villages usually have one and not more than one church. It is due to this general knowledge about the village as the larger discourse context that the referent of *the church* is contextually unique and thus mutually and unambiguously identifiable by both the speaker and the hearer.

Hence, we can say that contextually unique referents $R_{\text{cont.u}}$ are unambiguously identifiable because the discourse participants access a larger context based on general knowledge in which the referents are unique. Contextually unique referents can be defined as follows:

- (101) *Contextually unique referent* $R_{\text{cont.u}}$

A referent evoked by a referential expression is a contextually unique referent $R_{\text{cont.u}}$ if the discourse participants can access a larger context based on general knowledge in which the referent is the only (salient) referent of its kind. A contextually unique referent is therefore mutually and unambiguously identifiable, and is assigned to the same referential expression by the speaker and the hearer.

A clear-cut distinction between situational and contextual uniqueness is not always evident. Consider example (102), where the location of the discourse situation, the new flat of H, provides the context in which the referent of *kitchen* is unique and therefore mutually and unambiguously identifiable by the speaker and the hearer.

- (102) (Context: H is showing S her new flat.)
*Where is **the kitchen**?*

In (102) the larger context accessed by general knowledge can be argued to be the concept of ‘flat’ that usually has one and not more than one kitchen by default. However, in this case the larger

context essentially corresponds to the immediate discourse situation taking place inside of the flat in question. It is therefore not evident whether the referent of *the kitchen* in (102) is situationally unique because of specific knowledge from the immediate discourse situation, or contextually unique because of general knowledge, or both at the same time. This issue is also mentioned in [Himmelman \(1997: 39-40\)](#), who notes that in discourse, a number of different factors are usually involved in the definiteness of an expression. I follow this approach, assuming that it may not always be possible to determine a single function of a referential expression, i.e. that a referential expression may sometimes fulfill more than one function. Regarding example (102), we could thus say that the referent of *the kitchen* is both situationally and contextually unique.

In addition to the examples discussed so far, there are a number of constructions that involve definite marking and that can be subsumed under contextual uniqueness. One group of such constructions include a definite marker (e.g. a definite article) and some form of possessive or genitive marker.¹² One sub-type of such expressions can be accounted for in a straightforward way by [Löbner's](#) classification of concept types, as mentioned earlier in this section. The types of concepts that we often find with a definite marker and a possessive or genitive marker are in fact functional concepts, which inherently include a possessor argument and which are linked to that argument in an unambiguous way so that they can always be constructed as contextually unique based on general world or cultural knowledge.¹³ Examples of such functional expressions are given in (103) and (104).

- (103) a. *the **top** of the/a mountain*
 b. *the **mother** of the/a child*
 c. *the **frontpage** of the/a magazine*
- (104) a. *Lisa's **age***
 b. *Peter's **stomach***
 c. *her **name***

English only uses definite articles in constructions like the ones in example (103) but not in (104). However, in other languages, we may also find definite marker or articles in genitive and possessive constructions similar to the ones in (104). The important point is that the referents of such functional concepts in both (103) and (104) are contextually unique, the context being provided by the possessor argument.

¹²Hawkins (1978: 241-245) discusses such expressions as “associative clauses”.

¹³Note that [Löbner \(1985, 2011\)](#) assumes that the property of being a functional concept is part of the lexical meaning, which is a different way of saying that the properties are part of general world (and linguistic) knowledge in the sense of [Hawkins \(1978\)](#).

Interestingly, expressions similar to the ones in (103) also require the use of the definite article even if the noun in the definite expression is not a functional concept. Examples of such expressions are shown in (105):

- (105) a. *I usually had breakfast at **the corner** of a major intersection.*
(Poesio 1994: 283)
- b. *As you know, I didn't expect to be **the parent** of a hyperactive child.*
- c. *The baby's fully-developed hand wrapped itself around **the finger** of the surgeon.*
- d. *Alexander the Great was crossing the desert on his donkey. Suddenly **the leg** of the donkey buckled and it fell.*
(Barker 2004: 96-97)

In each of the examples in (105), the highlighted expression marked by the definite article is identified or anchored by the concept in the genitive modifier. For instance, *the corner* in (105a) is anchored by the referent of the expression *a major intersection*. Similarly to the functional concepts in (103) and (104), the concepts of 'corner', 'parent', 'finger', and 'leg' can be said to involve an additional possessor-like argument in their lexical properties. In contrast to functional concepts, however, the relation between the referent of the noun and the additional argument is not unambiguous. Löbner (1985, 2011) labels them relational concepts. Because their link or anchor to the referent of the possessor argument is not a one-to-one mapping, referents of relational concepts, as in (105) are not contextually unique in the strict sense. Moreover, the intersection in (105a) inherently and necessarily has more than one corner. Barker (2004: 96) makes this case by showing that (105c) is a felicitous description of a picture in which three fingers of the surgeon are visible, only one of which the baby's hand holds. This means that the immediate discourse context can entail the non-uniqueness of a referent, while the referring expression is marked as definite, i.e. as unambiguously identifiable.

In the same way as previous studies have argued for accommodation, one could say that contextual uniqueness can also be constructed by the hearer. Thus, in (105a), only one of the four corners of the intersection is relevant for the following discourse because it is the corner that the speaker had breakfast at. In the same way, in (105b) it is not relevant if, as expected by default, there is another parent other than the speaker;¹⁴ nor is it relevant in (105c) that the surgeon has other fingers that the baby did not touch, nor that the donkey had another leg in (105d). In other words, the referents related to these concepts can be constructed as contextually unique because the other (potentially) contextually salient referents of their kind are irrelevant for the following

¹⁴Note that this example also illustrates that referents can be unambiguously identifiable based on more than one condition or factor. Assuming that the child in (105b) has two parents, and assuming that the other parent is not present in the immediate discourse situation, then the speaker does correspond to the only referent of *the parent* within the immediate discourse situation, resulting in the referent also being situationally unique.

discourse. In contrast to the examples in which the hearer needed to anchor referents in larger contexts such as the village or the country of the discourse situation, in these cases, marking the referents as contextually unique requires the hearer to restrict the context further to only the relevant referent.

This subtype of contextually uniqueness can also be observed with concepts, e.g. sortal concepts in the terminology of Löbner, Löbner as shown in (106). The referent of *her bike* in (106a) is the only contextually salient referent of its kind, anchored to the referent of *Linda* and being contextually unique, without a one-to-one relation between Linda and the bike being necessary. The same holds for the referent of *the house* in (106b); it is the only contextually salient or relevant of its kind, even though it is imaginable that the neighbor owns more than one house. I thus treat referents of such types of expressions as contextually unique as well.

- (106) a. *This morning Linda had an accident and fell from **her bike**.*
b. ***The house** of my neighbor was set on fire.*

The last type of contextually unique referents are those whose expression involves superlatives that inherently include (contextual) uniqueness in their lexical properties, presupposing that there cannot be another referent that would satisfy the description of the referent in question. An example of a referent whose contextual uniqueness is based on a superlative expression is given in (107). Similarly to superlatives, certain types of quantifiers or other related expressions also include a uniqueness presupposition and thus make the referent contextually unique as well. Some examples of this latter type are shown in (108), serving as an illustration of the phenomenon rather than being an exhaustive list.

- (107) *Tina is **the smartest student** at the conference.*
- (108) a. *You are **the first visitors** to our new house.*
b. *I've got **the same problems** as you.*
c. ***The next water** is beyond those hills.*
(Lyons 1999: 10-11)

Related to contextually unique referents is a phenomenon discussed as “weak definites” in the literature (e.g. Aguilar-Guevara & Zwarts 2013, Carlson & Sussman 2005, Schwarz 2014, J. Zwarts 2014). Three examples of weak definite expressions are given in (109).

- (109) a. *Larry answered **the phone**.*
b. *Lucas had to go to **the hospital**.*
c. *Every morning, Fiona is reading **the newspaper**.*

What such weak definites in (109) have in common is that the referents assigned to such expressions are no longer necessarily contextually unique. For instance, the contexts of (109a) and (109b) can include more than a single contextually salient referent of *the phone* or *the hospital*. Larry could be sitting at a desk with more than one phone, and Lucas could be situated in a city with several hospitals. Nevertheless, the referring expressions *the phone* and *the hospital* include the definite article whose use is felicitous in such contexts. Also (109c) can be interpreted in a way that the speaker makes reference to more than one referent of *the newspaper*; the default interpretation is that Fiona reads a different newspaper every morning.

Carlson & Sussman (2005) use VP ellipsis as in (110) and (111) to show that weak definites are no regular definite expressions with a (contextually) unique referent and to test whether an expression is a weak or a strong definite one. *The store* being a weak definite in (110) as opposed to *the desk* in (111), the referent of *the store* does not have to be the same in the coordinated clauses, whereas it does in the case of *the desk*.

(110) *Fred went to the store, and Alice did, too.* (OK as different stores)
(Carlson & Sussman 2005: 72)

(111) *Fred went to the desk, and Alice did, too.* (must be the same desk)
(Carlson & Sussman 2005: 72)

Already Carlson & Sussman (2005) note that weak definites have similar properties to bare nouns in expressions such as *go to bed*, *be in prison*, or *attend school*. Building on this, Aguilar-Guevara & Zwarts (2010, 2013) argue for an analysis of weak definites as referring to a kind and not to a particular instance. These accounts thus treat weak definites as type of generic expressions.

J. Zwarts (2014) proposes a different analysis in which weak definites refer to roles in frames. Informally, he describes them as follows:

So, when we partition our world into certain temporal units (like days and weeks) or spatial units (like rooms, homes, buildings, towns, countries), or spatio-temporal combinations of these, then we tend to find entities fulfilling more or less the same constant functions across those units. For example, every day has a routine with meals, school, university, work, trains. In our spatial environment we will usually find radios, televisions, mirrors, calendars, and other object with stereotypical and constant functions. The idea is then that weak definites pick up those types of entities that are uniquely given in such general frames as roles or slots.
(J. Zwarts 2014: 274)

As described for contextually unique referents above, weak definites can be thought of as referring to contextually unique referents in conventionalized frames in the sense of temporal or spatial units. The referent of *the phone* in (109a) is thus evoked in a frame of ‘calling’ or ‘talking on

the phone’, which usually involves a single phone per person. This in turn makes the referent of *the phone* the single salient (or relevant) phone in this frame. Similarly, the referent of *the hospital* can be thought of as the role of the place of treatment within the frame of ‘accidents’, ‘illnesses’, or frames of other circumstances that require medical attention. Within this frame, the role of ‘hospital’ is unique by default, since hospitalization usually only involves a single hospital. The difference between weak definites and contextually unique referents is that weak definites refer to a unique role within a frame, making them a function between the frame and the realization of the referent, whereas contextually unique referents correspond to instances of realized referents within a given context.

Despite their similarity to contextually unique referents, weak definites are no typical definite referents and can be argued to have properties of generic referents. Both weak definites and generic expressions can but do not necessarily have to be expressed by a definite article. For instance in English, the use of the definite article with such referents is arguably construction-specific. It is required in *go to the store* but absent in *go to prison*. Moreover, the distribution of the definite article in such contexts across languages is not systematic either.¹⁵ Although an article can or may be required to be used in certain weak definite contexts, its use is construction-specific and cannot be considered to be a primary or necessary one. Therefore, weak definites will not be considered in the remainder of this study.

3.2.7 Bridging

Bridging referents can be thought of as a subtype of contextually unique referents but with similarities to anaphoric referents as well.¹⁶ Therefore, and because bridging referents have been discussed as a separate function in the literature, I treat them as a function here as well. Bridging occurs when a nominal expression introduces a new referent R_{brid} , which is nevertheless licensed as definite by a special relation to a previously mentioned referent R_{ante} . Three typical bridging examples are given in (112).

- (112) a. *John bought a book. **The author** is French.*
 b. *I just flew back from Colombia last night. **The flight** was exhausting.*
 c. *I saw an old building in the village. **The roof** had been severely damaged.*
 d. *Lisa went to a wedding yesterday. **The bride** had a beautiful dress.*

¹⁵Dryer (2014: 237) makes the following remark about the use of definite articles with weak definites: “I believe, however, on the basis of what I have observed about how definite articles are used in texts for other languages, that such noun phrases would rarely be marked as definite in other languages.”

¹⁶The term “bridging” goes back to Clark & Haviland (1977), but the phenomenon was already noted in Christophersen (1939), and was discussed as “associative anaphora” in Hawkins (1978), as “inferrables” in Prince (1981), as “accommodation” in Heim (1988), and as “indirect anaphora” in Erkü & Gundel (1987).

The bridging relation between *a book* and *the author* in (112a) is not a proper anaphoric relation because the referents are not identical, but they are nevertheless linked in an unambiguous way, and one could imagine R_{book} as an antecedent of R_{author} , with a more complex relation between the two referents than in the case of a proper anaphoric relation. This also holds for the relation in (112b), except that now the antecedent is a verbal expression. Also the relation between *an old building* and *the roof* in (112c) could be explained by a special anaphoric relation, although the bridging referent is a part of the antecedent. For the relation between *a wedding* and *the bride*, on the other hand, one may argue that the concept of ‘wedding’ evokes a context with a single bride, making the referent *the bride* somewhat more similar to a contextually unique referent.

Hawkins (1978) classifies such contexts as a single type, called associative anaphora, which differs from anaphoric as well as larger situation uses. Schwarz (2009), on the other hand, providing empirical evidence from German and Frisian, argues that bridging contexts do in fact not form a homogeneous category of contexts but should rather be understood as two special types of contextually unique and anaphoric referents. In this study, I follow Schwarz (2009) in making the distinction between two types of bridging, but I nevertheless group them together as bridging contexts because they do have in common that they involve a linguistically expressed antecedent in contrast to contextually unique referents. Moreover, even though Schwarz (2009) provides ample evidence for how the two types of bridging affect the use of articles in German and Frisian in a different way, these differences are rather tendencies than absolute differences in certain types of contexts. This suggests that not all bridging contexts can always be assigned categorically to one of the subtypes, that their classification may differ across languages or that there are different interpretations available. Still, the two subtypes will be relevant to the use of anaphoric (strong definite) and weak definite articles (cf. Sections section 5.2 and section 5.3), which is why the following paragraphs will present both types as defined by Schwarz (2009).

The first type of bridging is called part-whole bridging by Schwarz (2009), as typical contexts of this type involve part-whole relations like *building–roof* in (112c). I will refer to this function as unique bridging (u-bridging). U-bridging can be thought of as a type of contextual uniqueness, the difference being that contextually unique referents are constructed as unique in a given context without any linguistic expression of that context. U-bridging referents, on the other hand, occur with an antecedent which evokes the context in which the u-bridging referent is to be constructed as unique. As was argued in Section 3.2.6 for contextually unique referents, this unambiguous link or anchor is available to the speaker and the hearer based on general world or cultural knowledge. Frame semantics (cf. Fillmore 1975, 1976) is inherently suited to account for the definite status of a bridging referent:

[...] frames and scenes, in the mind of a person who has learned the associations between them, activate each other; and that furthermore frames are associated in memory with other

frames by virtue of their shared linguistic material, and that scenes are associated with other scenes by virtue of the entities or relations or substances in them, or their contexts of co-occurrence. (Fillmore 1975: 124)

Applied to the *building–roof* example in (112c), this means that when *a building* is uttered, its frame evokes the concept of ‘building’. Based on our world knowledge about the parts of buildings, we know that the evoked context will contain one single referent of the concept of ‘roof’. This links the referent R_{roof} to the referent R_{building} and requires the noun *roof* to occur with the definite article. The latter marks its referent as unambiguously identifiable, even though it is mentioned for the first time. The same holds for the *wedding–bride* example in (112d). The referent R_{wedding} evokes a context with a single bride, which makes the referent R_{bride} contextually unique, and which requires it to be marked by the definite article.

Like contextually unique referents, unique bridging referents are mutually and unambiguously identifiable because they are (constructed as) the only referent of their kind in a certain context. The difference between contextually unique and bridging referents only lies in how this larger context is evoked. For contextually unique referents this context is either implicit, as in *Where is the kitchen?*, or it is part of the expression of the referent itself, as in *the top of the mountain*. U-Bridging referents, on the other hand, are more similar to anaphoric referents in that they are linked to a preceding referent which provides the context in which the u-bridging referents are unique. We can thus them as follows:

(113) *Unique bridging referent* $R_{\text{u-bridge}}$

A referent evoked by a referential expression is a u-bridging referent $R_{\text{u-bridge}}$ if it is the only salient referent of its kind due to its unambiguous link to another previously mentioned referent from the discourse universe, evoking the context in which the u-bridging referent is unique. A u-bridging referent is therefore mutually and unambiguously identifiable, and is assigned to the same referential expression by the speaker and the hearer.

The second type of bridging is called relational anaphora by Schwarz (2009); I will refer to this referent type as relational bridging (rel-bridging). From the examples shown above, (112a) and (112b) have rel-bridging referents. In (112a), the first part of the utterance refers to a book, adding the referent of *a book* to the discourse universe. In the following part of the utterance, the referent of *the author* is again licensed as definite for its link to the already established referent of *a book*. The way this link or relation is established for rel-bridging referents differs from the uniqueness account of u-bridging referents.

Schwarz (2009) argues that the relation between the two referents is established because the *author* is a relational noun that comes with an inherent possessor argument, allowing for expressions like *the author of the book*. In contrast to u-bridging referents, rel-bridging referents are

not constructed as contextually unique. Instead, the relation from the concept of ‘author’ to the concept of ‘book’ allows for an anaphoric relation between the antecedent and the rel-bridging referent. This anchors the referent of *the author* to the referent of *a book*. In contrast to the *wedding–bride* relation, where the context evoked by *a wedding* constructs the bride as unique, no such large, abstract context is evoked in rel-bridging cases and an anaphoric relation between e.g. the referent of *a book* and the referent of *the author* is established. This holds for other product-producer relations and contexts like the one shown in (112b), where the rel-bridging referent of *the flight* can establish an anaphoric relation to the preceding referent expressed by the verb *flew*, which makes the rel-bridging unambiguously identifiable but, crucially, not contextually unique.

Therefore, rel-bridging referents are similar to anaphoric referents in that the relation to a preceding referent makes them unambiguously identifiable. They differ from anaphoric contexts in that the latter involve a shared identity between referents, while rel-bridging referents need to have special properties so that they inherently include this relation to another referent. One account would be to say that they have to be relational concepts with an inherent argument role. Another way of saying that is that the relation has to be given by general or world knowledge, for instance that there is an inherent relation between the event of ‘flying’ and the concept of ‘flight’, or between the concept of ‘author’ and the concept of ‘book’. Rel-bridging referents can be defined as follows:

(114) *Relational bridging referent* $R_{\text{rel-bridge}}$

A referent evoked by a referential expression is a rel-bridging referent $R_{\text{rel-bridge}}$ if it is the only salient referent of its kind due to its link to another previously mentioned referent from the discourse universe, establishing an anaphoric relation between the two referents. A rel-bridging referent is therefore mutually and unambiguously identifiable, and is assigned to the same referential expression by the speaker and the hearer.

As was mentioned before, in certain contexts both types of bridging may apply. Consider examples (115) and (116).

(115) *They’ve just got in from New York. **The plane** was five hours late.*
(Lyons 1999: 3)

(116) *John was driving down the street. **The steering wheel** was cold.*
(Schwarz 2009: 6)

In both cases, a larger context can be evoked in which the bridging referent can be constructed as unique, but both bridging referents of *the plane* and *the steering wheel* inherently include the concepts of ‘flying’ and ‘driving’, which would also allow them to establish an anaphoric relation

to the antecedent. For the purposes of the present paper, it is not important to be able to distinguish between u-bridging and rel-bridging in each instance, and in many examples it is sufficient to treat them as bridging referents. However, the difference between these two types will be relevant to the distribution of anaphoric (strong definite) and weak definite articles discussed in Sections section 5.2 and section 5.3 so that it is important to be aware of these differences.

3.2.8 Absolutely unique

Absolute uniqueness is another referential function in the definite domain. Absolutely unique referents are elements of the intersection of the speaker's and the hearer's mental spaces without necessarily being part of the current discourse universe. Figure 3.1 in Section 3.1 showed this for the referents R_{sun} of *sun* and R_{moon} of *moon*, which are absolutely unique in the default perception of our world. They are unambiguously identifiable independently from the discourse situation or even a larger context, as they are the only referents of their kinds. In English, for instance, the definite article is licensed with the expression of such referents even without any previous mention:

(117) *The moon is bright tonight.*

Of course, this is not to say that we could not distinguish between different moons when talking about astronomy. Here I only refer here to a handful of concepts that are usually, even though not necessarily, perceived as absolutely unique. The reason for distinguishing this type of definite referent is that definite article in the world's languages are often unable to occur in such contexts (cf. Section 5.1.4); hence, this referential function has to be distinguished in order to be treated separately.¹⁷ Absolutely unique referents entail that they are the only possible referent of their kind. Therefore, the existence of another more salient referent of its kind is logically excluded, which makes the absolutely unique referent unambiguously identifiable and hence definite.

(118) *Absolutely unique referent* $R_{\text{abs,u}}$

A referent evoked by a referential expression is an absolutely unique referent $R_{\text{abs,u}}$ if it is the only referent of its kind. Therefore, both the speaker and the hearer can unambiguously identify the referent and assign the same referent to the referential expression.

3.2.9 Specific

Often a specific referent is understood as a referent that the speaker is able to identify, while the hearer is not, as in the following example:

¹⁷A more fine-grained distinction of noun types is proposed in Löbner (e.g. 1985, 2011). He distinguishes between sortal (e.g. *stone, book, water*), individual (e.g. *moon, weather, Maria*), relational (e.g. *sister, leg, part*), and functional (*father, head, age*) nouns.

(119) *I saw a new neighbor move in yesterday.*

In (119), the speaker can unambiguously identify the referent and assign it to the expression of *a new neighbor*. However, because the hearer does not have sufficient information to identify the referent, it is marked by the indefinite article *a* in English.¹⁸

The concept of specificity is also used in contrast with nonspecificity, the latter of which involves a discourse referent that neither the speaker nor the hearer can identify and that does not have any existence requirement (e.g. C. L. Baker (1973), Karttunen (1976), Hawkins 1978: 203f, Anderson & Keenan (1985: 179), Lyons (1999: 58)). This leads to the commonly found three-way distinction based on the combination of speaker and hearer identifiability, displayed in Table 3.1 (von Heusinger 2002: 249).

Table 3.1: Traditional view of specificity and nonspecificity

identifiable by	definite specific	indefinite specific	indefinite nonspecific
speaker	+	+	–
hearer	+	–	–

However, von Heusinger (2002) argues that this description of specificity is not adequate; he shows that we find instances of specific referents which are not identifiable by either the speaker or by the hearer. For the purposes of the present study, I will follow von Heusinger (2002) in that the description of specificity as identifiability by speaker and by the hearer is not sufficient. Consider the following example:

(120) *Sarah wanted to talk to a colleague of mine, but I don't know which one.*
(Ionin 2006: 182)

In (120), the referent of *a colleague of mine* is not unambiguously identifiable by either the speaker or by the hearer. Nevertheless, we can assume that Sarah knows which *colleague* she wants to talk to. Thus, the speaker and the hearer know or can trust that there is a particular referent that *a colleague of mine* is linked to, even though neither of them can identify this referent. Rather, both can assume that Sarah can unambiguously identify the referent and assign it to the expression *a colleague of mine*. Therefore, the referent of *colleague of mine* is anchored or linked to the referent of Sarah. We can hence define a specific referent R_{spec} as follows:

¹⁸This type of specificity has been labeled “epistemic” specificity in the literature; see Farkas (1994, 2002), von Heusinger (2002, 2007), Kamp & Bende-Farkas (2019) for a more fine-grained distinction of different types of specificity.

(121) *Specific referent* R_{spec}

A referent evoked by a referential expression is a specific referent R_{spec} if both the speaker and the hearer know that the referential expression evokes a particular referent, while the hearer, or the speaker and the hearer, cannot unambiguously identify the referent.

Since a specific referent R_{spec} is not mutually and unambiguously identifiable, it is part of the indefinite domain. This is reflected in the use of the indefinite article in English in examples such as (119) and (120).

3.2.10 Nonspecific

Like specific referents, nonspecific referents are part of the indefinite domain. However, they differ from both definite and specific referents in that nonspecific referents do not correspond to any particular referent from their kind. An example of a nonspecific referent is given in (122):

(122) *Lea wants to buy a bike, but she doesn't know if she will find one.*

In contrast to specific referents, nonspecific referents are usually assumed to lack an existence presupposition. In other words, in examples such as (122), the speaker does not assume that a referent necessarily exists. In the present study, I do not base the definition of nonspecificity on the existence of the referent. A nonspecific referent is a single referent from its kind, which can be thought of as a “placeholder referent”,¹⁹ with no particular referent assigned to the expression used. Since there is no further information about the identity of the referent from the kind ‘bike’ that is linked to the expression *a bike*, the referring expression could be assigned to any of them, and it is either not important or not known which one it is. A nonspecific referent R_{nspec} can thus be defined as follows:

(123) *Nonspecific referent* R_{nspec}

A referent evoked by a referential expression is a nonspecific referent R_{nspec} if the referential expression does not evoke any particular referent from its kind. Therefore, neither the speaker nor the hearer can identify the referent, and no particular referent is assigned to the referential expression.

3.3 Defining the article types

This section defines the 10 different article types that can be found in the world’s languages and that cover various combinations of referential functions defined in the previous section. Detailed examples are provided in Chapters chapter 5 to chapter 7 for each article type.

¹⁹Similarly to definite expressions, we can imagine a group of referents to be nonspecific as well. In that case, the nonspecific referent would correspond to any subset with more than one referent from its kind.

3.3.1 Articles in the definite domain

The main function of definite articles is to mark the referent of the noun that they occur with as identifiable for all discourse participants. This is not to say that definite articles could not occur in other, additional contexts. However, other uses of definite articles are not relevant for their definition. For instance, they may or may not occur in generic contexts, but this cannot be a criterion to classify a marker as a definite article or not. Nor are definite articles required to encode all subtypes of definite referents.

Deictic referents can but do not need to be marked by definite articles. We find definite articles that do not originate from spatial demonstratives and that are systematically used in the relevant definite contexts without occurring in deictic contexts (cf. Section 5.1.3). A similar restriction concerns absolutely unique referents (cf. Section section 5.1.4); definite articles in the world's languages differ as to whether or not they can encode this function. Since these articles are otherwise systematically used to encode all remaining referential functions from the definite domain, I regard them as definite articles, which will be presented in detail in Section 5.1. Definite articles are thus defined as follows:

(124) *Definite article*

A definite article is an article that systematically marks anaphoric, recognitional, establishing, situationally unique, contextually unique, and bridging referents. It may also occur in other types of definite or generic contexts. It does not mark a referent as specific or nonspecific.

Since the ability of definite articles to encode recognitional referents is not systematically documented in grammars, I did not include the use of a marker with recognitional referents as a criterion for definite articles in the sample for practical reasons. From a theoretical perspective, however, definite articles are assumed to encode recognitional referents. The same holds for establishing referents. Following [Himmelmann \(1997: 93-101\)](#), definite articles are expected to encode both recognitional and establishing referents. He argues that these two functions play a central role in the development of definite articles from demonstratives. This means that the sample contains definite articles for which a recognitional or establishing use is not attested. For the definite articles in the sample, the criteria are thus that they need to systematically mark (at least) anaphoric, situationally unique, and contextually unique referents from the definite domain.²⁰

²⁰Examples of each article in the sample are provided in the online appendix; for each definite article, anaphoric and contextually unique (and if possible, situationally unique and establishing) uses are shown. In a few cases, bridging uses are shown instead of contextually unique ones. The same holds for examples illustrating the use of inclusive-specific and referential articles in the definite domain.

Anaphoric articles represent another type of articles from the definite domain. They correspond to what Schwarz (2009) introduced as strong definite articles and will be discussed in detail in Section 5.2. We can define anaphoric articles as follows:

(125) *Anaphoric article*

An anaphoric article is an article that systematically marks anaphoric referents and may also mark recognitional, establishing, rel-bridging, and situationally unique referents. It does not mark deictic, contextually unique, or u-bridging referents. It does not occur in specific or nonspecific contexts either.

Weak definite articles, as introduced by Schwarz (2009), are in a way the counterpart of anaphoric (or strong definite) articles and will be discussed in Section 5.3. They only encode semantic definites. A definition of weak definite articles is given in (126).

(126) *Weak definite article*

A weak definite article is an article that systematically marks situationally unique, contextually unique and u-bridging referents. It does not mark deictic, anaphoric, recognitional, establishing, or rel-bridging referents. It does not occur in specific or nonspecific contexts either.

The last article type from the definite domain is the recognitional article, defined in (127) and described in detail in Section 5.4.

(127) *Recognitional article*

A recognitional article is an article that systematically marks recognitional referents. It does not have other referential functions.

3.3.2 Articles in the indefinite domain

In the indefinite domain, we find three types of articles: indefinite, exclusive-specific, and non-specific articles. The probably most prominent type of articles from the indefinite domain is the indefinite article, discussed in Section 6.1. Indefinite articles can be defined as follows:

(128) *Indefinite article*

An indefinite article is an article that systematically marks specific and nonspecific referents. It may be able occur in generic contexts, and it does not mark referents as definite.

Indefinite articles are semantically vague between coding specific and nonspecific reference, resulting in the ambiguity between specificity and nonspecificity in case the context does not resolve this ambiguity otherwise. Based on the notion of colexification proposed in François & Vanhove

(2008), Hartmann et al. (2014) introduce the notion of coexpression. Regarding indefinite articles, we can say that they coexpress specific and nonspecific referents.

Another type of articles found in the indefinite domain is what I call an exclusive-specific article. They will be presented in Section 6.3. Exclusive-specific articles correspond to what is usually referred to as specific article. I make the distinction between exclusive-specific and inclusive-specific because we find two types of specific articles in the languages of the world, as will be shown later. Exclusive-specific articles are defined as follows:

(129) *Exclusive-specific article*

An exclusive-specific article is an article that systematically marks specific referents. It does neither occur with definite nor with nonspecific referents.

The nonspecific article can be viewed as the counterpart of the exclusive-specific article within the indefinite domain. Nonspecific articles will be discussed in Section 6.4. Their definition is given in (130).

(130) *Nonspecific article*

A nonspecific article is an article that systematically marks nonspecific referents. It does neither occur with definite nor with specific referents.

3.3.3 Domain-crossing articles

There are three more types of articles in the world's languages that belong to neither the definite nor the indefinite domain; they encode referential functions from both domains. These article types are (weak) inclusive-specific and referential articles. In contrast to exclusive-specific articles, inclusive-specific articles include the coding of definite referents. They will be presented in detail in Section 7.1 and they can be defined as follows:

(131) *Inclusive-specific article*

An inclusive-specific article is an article that systematically occurs with anaphoric, recognitional, establishing, situationally unique, contextually unique, bridging, and specific referents. It may also occur with other types of definite referents and generic referents. It does not occur with nonspecific referents.

Again, due to the lack of examples of recognitional, bridging, and establishing referents in many descriptions, I rely on the systematic coding of anaphoric, situationally unique, and contextually unique referents as criteria for the article's coverage of the definite domain. As was mentioned for indefinite articles, expressing different referent types should be understood as coexpression, i.e. inclusive-specific articles do not disambiguate between the referential functions that they encode.

By analogy to weak definite articles, we need to distinguished what can be called a weak inclusive-specific article. The use of weak inclusive-specific articles in the definite domain is restricted to semantic definite contexts; they can be defined as follows:

(132) *Weak inclusive-specific article*

A weak inclusive-specific article is an article that systematically marks situationally unique, contextually unique, bridging, and specific referents. It does not mark deictic, anaphoric, recognitional, establishing, or rel-bridging referents. It does not occur with nonspecific referents either.

Referential articles have the broadest referential functions of all article types. As will be shown in Section 7.2, they no longer distinguish between the main referent types of definite, specific, and nonspecific referents. In contrast to other article types, referential articles do not mark the referential function of the noun that they occur with; rather, they indicate that the noun is to be interpreted as a referential expression. There are languages which contrast referential articles with predicate markers. Such cases nicely show that a referential article does not only function as a general referential marker but also as an argument marker from a syntactic point of view. In these systems, the referential article is then opposed to the predicate marker, which is used when lexical referential expressions should be interpreted as predicates, i.e. as non-referential. Referential articles are defined as follows:

(133) *Referential article*

A referential article is an article that systematically occurs with anaphoric, recognitional, establishing, situationally unique, contextually unique, bridging, specific, and nonspecific referents. It may also occur in other types of definite and generic contexts.

Like definite and inclusive-specific articles, I assume that referential articles mark all types of definite referential functions listed in (133). Since recognitional, bridging, and establishing uses are not always included in the descriptions of the languages in the sample, I use the occurrence with anaphoric, situationally unique, and contextually unique referents as criteria for the use of referential articles in the definite domain.

3.4 Referent types, article types, and referential hierarchies

This section is a synthesis of the previous sections in that it describes the relation between the different referential functions and article types. To address this question in more detail, I will also briefly present referential hierarchies and scales proposed in earlier studies and compare them to the distinctions made here.

3.4.1 Referent types and article types distinguished in the present study

Table 3.2 summarizes the referent types that are relevant for the purposes of this study as defined in Section 3.2.

Table 3.2: Overview of the referential functions

R_{deictic}	A deictic referent is identifiable based on its unambiguous link to an object in the discourse situation. <i>Do you see the house over there?</i>
R_{anaph}	An anaphoric referent is identifiable based on its shared identity with a previously mentioned referent. <i>A: Did you like the movie last night? B: The movie was great!</i>
R_{recog}	A recognitional referent is identifiable based on experience or knowledge previously shared by the speaker and the hearer. <i>What happened to the cat (we used to have)?</i>
R_{establ}	An establishing referent is marked as identifiable and thus constructed as such. <i>Did you hear the news? They are going to close the museum.</i>
$R_{\text{rel-bridge}}$	A rel-bridging referent is identifiable based on its relation to another previously mentioned referent, allowing for an anaphoric relation between the two referents. <i>Eva bought a book. The author is French.</i>
$R_{\text{sit.unique}}$	A situationally unique referent is identifiable because it is the only salient referent of its kind in the immediate discourse situation. <i>Beware of the dog.</i>
$R_{\text{cont.unique}}$	A contextually unique referent is identifiable because it is constructed as the only salient referent of its kind in a larger context. <i>What's the best way to the center?</i>
$R_{\text{u-bridge}}$	A u-bridging referent is identifiable based on its relation to another previously mentioned referent, which allows to construct the u-bridging referent as contextually unique. <i>I saw an old house in the village. The roof had been severely damaged.</i>
$R_{\text{abs.unique}}$	An absolutely unique referent is identifiable because it is the only referent of its kind. <i>The Earth is round.</i>
R_{spec}	A specific referent is not identifiable but linked to a particular referent of its kind. <i>I met a strange new neighbour yesterday.</i>
R_{nonspec}	A nonspecific referent corresponds to a single, but no particular referent of its kind. <i>Do you have a pen? Any pen will do.</i>

Deictic and absolutely unique referents are included in this overview rather because of their prominence in discussions of definiteness in the literature. As explained in Section 3.3.1, I do not view those two referential functions as relevant for the definition of article types in the definite domain.

Therefore, they are grayed out in Table 3.2, leaving 7 relevant functions in the definite domain and 2 functions in the indefinite domain. Building on that, Table 3.3 shows how the 10 article types defined in Section 3.3 can be mapped onto their referential functions.

Table 3.3: Mapping the article types to their referential functions

article types	referent types							indefinite domain	
	definite domain							R_{spec}	R_{nspec}
	R_{anaph}	R_{recog}	R_{establ}	$R_{\text{rel-bridge}}$	$R_{\text{sit.unique}}$	$R_{\text{cont.unique}}$	$R_{\text{u-bridge}}$		
DEF	✓	✓	✓	✓	✓	✓	✓		
ANA	✓	(✓)	(✓)	(✓)	(✓)				
DEF _{WEAK}					✓	✓	✓		
RECOG		✓	(✓)						
INDEF								✓	✓
EXSPEC								✓	
NSPEC									✓
INSPEC	✓	✓	✓	✓	✓	✓	✓	✓	
INSPEC _{WEAK}					✓	✓	✓	✓	
REF	✓	✓	✓	✓	✓	✓	✓	✓	✓

There are 4 types of articles in the definite domain. The broadest type of articles in the definite domain is the definite article (DEF), covering the entire definite domain. Anaphoric articles (ANA), whose main function it is to encode anaphoric referents, can also be used to encode recognitional, establishing, and rel-bridging functions. In certain cases, their use even extends to situational uniqueness. Importantly, they are not used for contextual uniqueness or u-bridging, i.e. prototypical semantic definite contexts. Weak definite articles (DEF_{WEAK}) are the counterpart of anaphoric articles within the definite domain. They only encode the uniqueness-based definite functions, i.e. situationally and contextually unique as well as u-bridging referents. They do not occur in the other, pragmatic definite contexts. The last article type in the definite domain is the recognitional (RECOG) article, which is used to encode recognitional referents. This type is very rare across languages; based on the five examples in the sample, it appears that recognitional articles are generally able to mark establishing referents as well. Therefore, their occurrence with the latter referent type is shown in brackets in Table 3.3. In the indefinite domain, we need to distinguish indefinite articles (INDEF), encoding both specific and nonspecific referents, from exclusive-specific (EXSPEC) and nonspecific NSPEC articles, encoding specific and nonspecific referents, respectively. Finally, three types of domain-crossing articles are attested. Inclusive-specific

article (INSPEC) expresses all functions in the definite domain and also encode specific referents, while weak inclusive-specific articles (INSPEC_{WEAK}) do not occur with pragmatic definites. The referential article (REF) covers the entire definite as well as the indefinite domain, i.e. all the relevant functions distinguished in Table 3.3.

3.4.2 Referential scales

In order to capture the relation between the different referential functions, a number of hierarchies or other structures have been proposed in the literature. This section will introduce the most relevant ones and compare them to the referential distinctions proposed in the present study. I will return to the question of referential hierarchies in Section 8.4, building on the findings of the present study the present to propose a refined referential hierarchy.

The notable hierarchies that span both the definite and indefinite domain are the definiteness or referentiality hierarchy (Aissen 2003, Comrie 1989: 132-135, Croft 2003: 128-132, von Heusinger 2008, among others), and the reference hierarchy in Dryer (2014), going back to Givón (1978).²¹ What they have in common is that they structure and distinguish referential functions by mapping linguistic forms onto these functions, similarly to the arrangement in Table 3.3, where the specific referent is placed between the definite referents and the nonspecific referent because of how these referent types are expressed by articles. The traditional definiteness hierarchy is given in (134).

- (134) **Definiteness hierarchy** (Croft 2003: 132)
 definite > specific > nonspecific

Its extension, introduced as the referentiality scale by von Heusinger (2008: 5), includes the person and referentiality hierarchies from Croft (2003: 130) and adds non-argumental expressions to the scale. The extended referentiality scale is shown in (135).

- (135) **Referentiality scale** (von Heusinger 2008: 5)
 personal pronoun > proper noun > definite NP > indefinite specific NP > indefinite non-specific NP > non-argumental

The definiteness hierarchy shown in (134) has often been argued for in the context of differential object marking, especially to account for the patterns in Spanish, Turkish, and Persian.²² In

²¹A similar hierarchy is the Givenness hierarchy proposed in Gundel et al. (1993), distinguishing six givenness statuses, which correspond to “six cognitive statuses relevant to the form of referring expressions in natural language discourse” (Gundel et al. 1993: 275). I will not discuss this hierarchy in detail here because it includes a combination of referential and information-structural distinctions. Also, the referential distinctions used for the Givenness hierarchy largely correspond to the ones made in the other hierarchies.

²²Note that Croft (2003) refers to Comrie (1989: 132-135), who incorrectly states that accusative marking in Turkish is reserved to definite referents. As many have argued, Turkish accusative marking is used for definite and specific referents (e.g. Enç 1991, Hedberg et al. 2009, von Heusinger & Kornfilt 2005).

both Persian and Turkish, accusative marking is used for definite and specific referents (Enç 1991). Spanish marks patients in transitive contexts if they are animate referents and definite or specific; but in contrast to Persian or Turkish, nonspecific animate referents can be marked in Spanish under certain circumstances; only non-argumental (or non-referential) animate expressions do not receive the object marker (von Heusinger 2008). Glossing over the complex details of differential object marking, we can say that it reflects and confirms the hierarchical structure between referent types that results from their expression by articles. For instance, as shown in Table 3.3, inclusive-specific articles coexpress various types of definite referents and specific referents, while indefinite articles encode both specific and nonspecific referents. An article that would coexpress a referent type from the definite domain and nonspecific referents without expressing specific referents does not seem to be attested, warranting the definite referent types to be located at one end of the hierarchy (or scale), and the nonspecific referent on the other, with the specific referent in an intermediate position.

The extended referentiality scale in (135) additionally includes personal pronouns and proper nouns at the definite end of the scale, both of which are not relevant for the present study. However, it is no coincidence that the two types of expressions at the most referential end of the scale are not of major concern for articles. Because personal pronouns and proper nouns are strongly referential and definite, there is little functional pressure to additionally mark them as such by articles. This is why articles usually do not occur with personal pronouns (although they can in some languages), and this is also why their use of proper nouns differs across languages, as is briefly shown for definite articles in Section 5.1.5. On the other end of the referentiality scale in (135), von Heusinger includes non-argumental expressions. These correspond to what I call non-referential expressions. Including those delimits referential articles, which encode all relevant referential functions. Crucially, as will be show in Section 7.2, they do not occur with non-argumental or non-referential expressions, which also follows from the referentiality scale in (135).

Another important hierarchy is the reference hierarchy proposed by Dryer (2014: 235) shown in (136). This hierarchy is explicitly built on the basis of the crosslinguistic properties of articles observed in the sample of the two WALS chapters on definite and indefinite articles (Dryer 2013b,c).

- (136) **Reference hierarchy (Dryer 2014):**
 anaphoric definite >
 non-anaphoric definite >
 pragmatically specific indefinite >
 semantically specific indefinite >
 semantically nonspecific indefinite

The distinction of referential functions in this hierarchy is very similar to the distinction of referent types proposed in the present study. Since this hierarchy is also built on the distribution of articles, it does not include deictic definites either. Anaphoric referents are located at the top of the hierarchy, followed by non-anaphoric definites. Dryer (2014: 236) treats referents as non-anaphoric definites if they are identifiable “based only on shared knowledge of the speaker and hearer”. As many previous studies have shown and as was discussed in detail in Section 3.2, there are several types that need to be distinguished; Dryer (2014) only mentions absolutely unique concepts such as *sun* and *moon*. The main idea of this distinction probably is the distinction between weak and strong definites, discussed in detail in the following section. There I will return to anaphoric vs. non-anaphoric referents in the sense of Dryer as well.

The next referential function that Dryer (2014) distinguishes are “pragmatically specific indefinites”. These fall into the group of expressions that I treat as establishing definites and what have been discussed as cataphoric or indefinite uses of demonstratives in the literature, e.g. *I went to **this movie** last night, ...* (Dryer 2014: 237). Here, *this movie* is not yet identifiable to the hearer. Using the demonstrative, however, the speaker signals to the hearer that the referent is highly relevant for the following discourse segment and soon identifiable by the hearer as well. Dryer (2014) treats them as indefinites following the tradition of Gernsbacher & Shroyer (1989), Gundel et al. (1993), and Prince (1981). As explained in Section 3.2, I treat such referent types as belonging to the definite domain, following Hawkins (1978) and Himmelmann (1997), who call them “unfamiliar” uses of definites. Moreover, as will be shown in Chapter 5, establishing referents pattern with other pragmatic definite referent types, i.e. with anaphoric and recognitional referents. That is, establishing referents can usually be expressed by anaphoric articles but not by weak definite articles. Therefore, establishing referents as defined in the present study should be located more closely to anaphoric definites on the hierarchy. The difference between establishing referents as defined here and pragmatically specific indefinites in Dryer (2014) is that Dryer includes expressions that are marked as non-identifiable (i.e. indefinite), which warrants their lower status on the hierarchy. In the present study, establishing referents only include those referents that are not yet identifiable but marked as such by the speaker, given that the relevant information to establish identifiability is provided in the immediately following discourse segment. In other words, I do not include referents that are not marked as definite in such contexts. Thus, the slightly different definition of establishing referents vs. pragmatically specific indefinites, even though they overlap to a certain extent, is the reason why they are located in different positions relative to other referent types in the present study and in Dryer (2014).

The lowest two functions on the hierarchy are the least definite ones: (semantically) specific and (semantically) nonspecific referents. Those two types correspond to the traditional distinction between specific and nonspecific referents within the indefinite domain as defined in Section 3.3.1.

Dryer's hierarchy shown in (136) is better understood as a scale and can be read as a semantic map in the traditional sense. The referential functions distinguished are ordered based on their coding by articles in the languages of the world, predicting that a single article can co-express different (also more than two) functions, but only as long as their are neighbouring functions in the structure:

The primary factor defining the ordering of these types on the hierarchy is the following: if a language has an article that is used for more than one type of noun phrase on the hierarchy, then the set of types it is used with will be a set that is contiguous on the hierarchy.

(Dryer 2014: 235)

This is a relevant prediction of possible types of articles, and I will return to Dryer's hierarchy in more detail in Chapter 8, concerned with the crosslinguistic distribution and availability of article types. Briefly put, the article types distinguished in the present study, representing all the articles types attested to the best of my knowledge, largely correspond to the types that fall out from Dryer's reference hierarchy or the referentiality hierarchy shown in (135). However, as was mentioned above, if pragmatically specific indefinites in the sense of establishing referents are to be included, they would need to be located between anaphoric and non-anaphoric definites in Dryer's hierarchy, given that they are expressed by anaphoric rather than weak definite articles. In addition, the existence of recognitional articles does not follow from Dryer (2014)'s hierarchy (nor from the other hierarchies). As was shown in Table 3.3, they need to be integrated between anaphoric and situationally unique referents together with establishing referents because, again, anaphoric articles but not weak definite articles can express recognitional referents (cf. Section 5.2.4 for a brief overview of anaphoric articles encoding recognitional and establishing referents). I will return to referential hierarchies in Section 8.4, proposing a refined hierarchy based on the ones presented here and the findings of the present study.

3.4.3 Pragmatic (strong) and semantic (weak) definiteness

This section will provide an introduction to the distinction between pragmatic (strong) and semantic (weak) definites first made by Löbner (1985), and the distinction between strong and weak definite articles proposed in Schwarz (2009).

The structure between different referential functions in the definite domain is made explicit in Löbner (1985), who makes the influential proposal to distinguish between pragmatic and semantic definites. This distinction has been adopted in many studies since then, including the present study. As we will see below, the distinction of pragmatic and semantic definites also largely corresponds to the functional domains of strong and weak definite articles in the sense of Schwarz (2009).

Pragmatic definites are those definite expressions whose unambiguous identifiability only arises within the context of the discourse situation. Pragmatic definites include situational uses in the sense of Hawkins (1978); Himmelmann (1997) also treats recognitional and establishing definites as pragmatic definites. With regard to the referential functions introduced in Section 3.2, we can group together deictic, anaphoric, recognitional, establishing, rel-bridging, and situationally unique referents as pragmatic definites.²³ The important property that these referential functions share is that their mutual and unambiguous identifiability depends on the (immediate) context, i.e. the discourse situation.

In opposition to pragmatic definites, Löbner (1985) groups together certain other functions as semantic definites. Semantic definites do not rely on the discourse situation but are mutually and unambiguously identifiable only due to the type of expression. According to Himmelmann (1997), semantic definites comprise abstract situational and associative anaphoric uses. Abstract situational uses correspond to contextually unique and absolutely unique referents as defined in Section 3.2. Associative anaphoric uses are bridging contexts. In terms of the referential functions distinguished in this study, semantic definites include contextually unique, u-bridging, and absolutely unique referents. As I will show in Section 5.1.4, the coding of absolutely unique referents are not relevant for the use of articles. Himmelmann (1997) uses the distinction of pragmatic and semantic definites to distinguish definite articles from demonstratives. The latter can only be used to encode definite referents that are pragmatic definites, and semantic definites are exclusively coded by the definite article (cf. Section 4.1).

Löbner (2011) offers a fine-grained distinction between different types of definite referential expressions on the basis of different combinations of concept types (going back to Löbner 1985) and several definite functions. He proposes the scale of uniqueness shown in Figure 3.2, arranging different types of expressions on a scale from pragmatic to semantic uniqueness.

²³Although the data from the articles discussed in the present study suggests that situationally unique uses can also pattern with semantic definites, building the bridge between pragmatic and semantic definiteness.

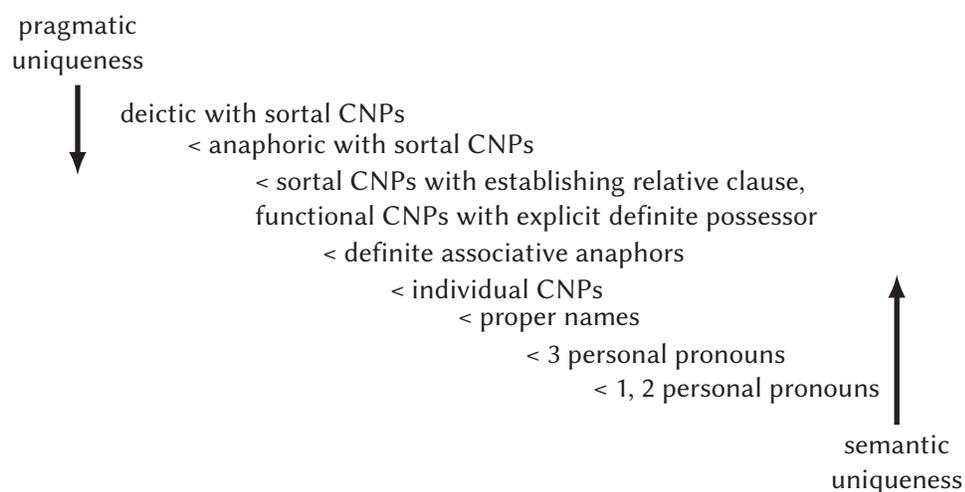


Figure 3.2: Scale of uniqueness (Löbner 2011: 320)

Deictic, anaphoric, and establishing uses of sortal concepts (e.g. *stone*, *book*, *water*) are pragmatic definite uses. Deictic and anaphoric uses, according to Löbner (2011), can always be encoded by demonstratives, which is why they are located at the extreme end of pragmatic definites. In Section 5.1.3, I take this up and argue that the expression of deictic referents plays no role in the definition of articles, which reflects the basic intuition that they are prototypically expressed by demonstratives, and definite article can but do not have to express deictic referents.

Even though establishing uses of sortal concepts are classified as pragmatic definites as well, they are positioned more towards the middle of the scale. This is because establishing uses can be expressed in the same way as semantically definite expressions in some languages according to Löbner (2011). As will be shown in Chapter 5, the data from the present study suggests that establishing referents are rather expressed by anaphoric articles than weak definite articles. On the other end of the scale shown in Figure 3.2, we find personal pronouns, proper names, and individual sortal concepts as semantic definites. The closer to the semantic definite extreme of the scale expressions are, the more inherently definite they are and the less likely they are to require any additional definite marking (Löbner 2011: 320), which was already mentioned in the previous section discussing referential hierarchies.

Based on this distinction between pragmatic (strong) and semantic (weak) definite referent types, Schwarz (2009) argues for a distinction of two types of definite articles, which are used to express pragmatic and semantic definite referents, respectively. The two types are therefore called strong and weak definite articles. Since its first proposal in Schwarz (2009), this distinction was taken up in various other studies (e.g. Arkoh & Matthewson 2013, Ingason 2016, Irani 2019, Jenks 2018, Ortman 2014, Schwarz 2013, 2019, Šereikaitė 2019).

Strong definite articles are used in anaphoric contexts only, where the referent is “familiar”. In contexts in which the referent is definite on the basis of uniqueness, as is the case with contextually unique referents, such a strong definite article cannot be used. Strong definite articles will be discussed in detail as anaphoric articles in Section 5.2.

In addition, a language may also have a separate weak definite article that is used with semantic definites, i.e. in uniqueness-based context only. Such weak definite articles are crosslinguistically very rare; they were proposed for only a handful of languages in the previous literature, which appears to be an exhaustive list. I will present and review weak definite articles in Section 5.3. The other logical possibility of a strong–weak distinction is that only one of the two groups of definites, pragmatic or semantic, are encoded by an article, while the other one is left unmarked. This pattern was noted for a number of languages (cf. various contributions in [Aguilar-Guevara et al. 2019](#)), which only have a strong definite article encoding pragmatic definite referents, whereas semantic definites are left unmarked. In the terminology of the present study, this corresponds to an article system with a single anaphoric article. Languages with such a system are indeed well represented in the sample and especially frequent in Australia. Interestingly, the reverse pattern is not attested, i.e. so far, we do not know of a language with a weak definite article but no anaphoric article ([Jenks 2018](#), [Schwarz 2019](#)). This issue is revisited in more detail in Sections 5.3 and 9.1.1.

To sum up, the present study generally follows the referential distinction made in the previous literature, combining the insights from [Hawkins \(1978\)](#), [Himmelmann \(1996\)](#), [Löbner \(1985\)](#), and [Schwarz \(2009\)](#), allowing for a fine-grained distinction of referential functions. This in turn serves as the basis for the distinction of the types of articles crosslinguistically attested.

3.5 Summary

This chapter dealt with the definition of different article types based on their referential functions. Section 3.1 introduced the relevant concepts related to referentiality. Section 3.2 then defined 11 referential functions, 9 of which are important to define different types of articles. These are: anaphoric, recognitional, establishing, rel-bridging, situationally unique, contextually unique, u-bridging, specific, and nonspecific. In Section 3.3, those functions were used to define the 10 attested types of articles: definite, anaphoric, weak definite, recognitional, exclusive-specific, nonspecific, indefinite, inclusive-specific, weak inclusive-specific, and referential articles. Section 3.4 discussed the relation between the article types and the referential functions expressed, comparing the referential distinctions made in this chapter to other distinctions and referential hierarchies from the literature, most notably to the distinctions between semantic and pragmatic definiteness.

Chapter 4

The distinction between articles and related markers

It is difficult to delimit certain types of articles from other similar or related markers, especially if they have a diachronic connection to articles. The four major cases that I will discuss in this chapter concern demonstratives and the articles in the definite domain (Section 4.1), possessives and definite articles (Section 4.2), the numeral ‘one’ and the articles in the indefinite domain (Section 4.3), as well as negative polarity items (NPIs) and nonspecific articles (Section 4.4).

The difficulty to distinguish between definite articles and demonstratives or between indefinite articles and the numeral ‘one’ is due to the grammaticalization from demonstratives and the numeral ‘one’ to definite and exclusive-specific or indefinite articles, respectively. Because this development is a gradual process, and because the functions of these markers overlap, it is important to clarify in more detail where the cut-off point between the source elements and the articles lies. Adnominal possessive markers represent another potential source for definite articles or markers of definiteness more generally, which is why I will briefly discuss the differences between possessive markers that appear to come close to markers of definiteness and definite articles. I will argue that the functional overlap can simply be accounted for by the inherent functions of the possessives, but that they do not provide strong evidence for a development from possessives to definite articles. The similarity between nonspecific articles and NPIs is not due to a diachronic relation. However, I will show that both types of markers essentially serve the same functions and that the difference between them lies in the systematicity of their use; nonspecific articles are required to mark nonspecific referents, NPIs can do so but are not required in the relevant contexts.

4.1 Demonstratives and articles in the definite domain

It is widely recognized that most definite articles in the world's languages originate from demonstratives. This grammaticalization path was already proposed by Christophersen (1939) and Greenberg (1978: 246), and was revisited in various studies since then (Diessel 1999, Leiss 2000, Lyons 1999, De Mulder & Carlier 2011). The grammaticalization path includes the development from the demonstrative to the definite article and the development from the latter to the inclusive-specific article following Greenberg (1978: 246-247) and De Mulder & Carlier (2011), as is shown in Figure 4.1:

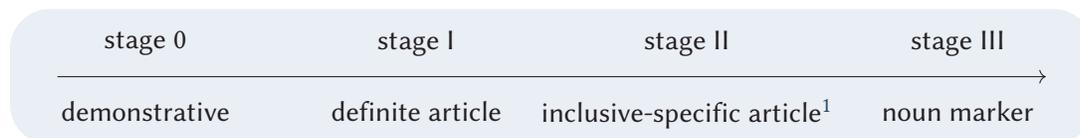


Figure 4.1: Grammaticalization path from demonstrative to general noun marker

The relevant part is the development between stage 0 (demonstratives) and stage I (definite articles) and the distinction of the properties of demonstratives on the one hand and definite articles on the other. In addition, even though not being part of the general grammaticalization path shown in Figure 4.1, anaphoric, weak definite, and recognitional articles need to be distinguished from the other two types of markers as well.

The demonstrative is understood as a deictic element, typically used for spatial deixis, indicating the relative position of an object with respect to the discourse situation with the aim to draw the hearer's attention towards the referent of this object. This use is called exophoric (cf. Diessel 1999: 94-95). Demonstratives with a primarily exophoric function can often be extended to anaphoric functions (Diessel 1999: 96). This use is also called endophoric (cf. Diessel 1999: 95-105). Exophoric and endophoric uses are thus two different types of deixis. For the sake of clarity, I use the term "deixis" only to refer to spatial deixis and treat exophoric uses as "anaphora". To be precise, deixis in this sense refers to physical perception and attention drawing. The prototypical deictic use of demonstratives can usually be accompanied by a pointing gesture (Diessel 2013: 243). This link between pointing gestures and demonstratives is well known in the linguistic literature (e.g. Brugmann 1904, Bühler 1934, H. H. Clark 1996, Eriksson 2009, Levinson 2004) and was often tied to the spatial deictic function of demonstratives. Diessel (2013: 243) points out that the function of such demonstratives is not only spatial location, but more importantly, drawing the hearer's attention towards an object in the discourse situation. He notes:

But a deictic pointing gesture is not just a guidepost for spatial orientation, it also serves to create what psychologists call a joint focus of attention (cf. Butterworth 1998, Eilan et al. 2005,

¹Greenberg (1978: 247) uses the label of "non-generic article" to refer to inclusive-specific articles.

Tomasello 1999). [...] Deictic pointing is the most basic communication device that people of all cultures use to establish or manipulate joint attention (cf. Kita 2003). (Diessel 2013: 243)

Hence, I will take the expression of joint focus of attention as a basic communicative function that is universal in human language. This is supported by the fact that demonstratives are generally taken as a universal part of speech (e.g. Evans & Levinson 2009; Diessel 2013: 245). This point will be relevant later for the argument to treat (certain types of) anaphoric and recognitional markers as articles and not as demonstratives.

As was discussed in detail in Section 3.4, various previous studies have pointed out that there is structure in the relation between different definite referential functions, and that they form a scale or continuum from the most pragmatic ones to the most semantic ones. It was often noted that the use of demonstratives is restricted to pragmatically definite referents, while definite articles typically occur in semantic definite contexts (Hawkins 1978, Himmelmann 1997, Löbner 1998). Pragmatic definite uses are those that do not rely on the inherent lexical meaning of the referent but on the immediate discourse situation. In opposition to those, semantic definites are not unambiguously identifiable based on the immediate discourse situation but based on inherent properties of the concept that the referent instantiates. Using this general distinction, Himmelmann (1997) provides an overview of the relevant pragmatic and semantic definite uses to distinguish between demonstratives and definite articles. These are shown in Table 4.1, which is adapted from Himmelmann (1997: 191):

Table 4.1: The functional domain of demonstratives and definite articles (Himmelmann 1997: 191)

	referential function	demonstrative	definite article
pragmatic definiteness	situational (= deictic & sit. unique)	✓	✓
	discourse deictic	✓	✓
	anaphoric	✓	✓
	anamnesic (= recognitional & establishing)	✓	✓
semantic definiteness	abstract situational (= cont. & abs. unique)	✗	✓
	associative anaphoric (= bridging)	✗	✓

In Table 4.1, deictic and situationally unique referents are grouped together as situational uses. The next type of contexts distinguished, the discourse deictic use, goes beyond the scope of the present study. Suffice it to say that discourse deictic uses of demonstratives refer back to a proposition or event mentioned earlier in the discourse (cf. Diessel 1999: 100-105, Himmelmann 1997: 84). The remaining two pragmatic definite uses shown in Table 4.1 are anaphoric and anamnesic uses. The latter corresponds to recognitional and establishing functions. The two semantic

definite functions are so-called abstract situational and associative anaphoric uses, which correspond to contextually unique and bridging functions as defined in the present study, respectively. This provides us with the necessary tools or contexts to distinguish between definite articles and demonstratives, if we restrict bridging contexts to u-bridging. If a marker can and has to be used with contextually unique and u-bridging referents expressed by nouns, then we can treat it as a definite article. If the marker does not (systematically) occur with those types of referents but only with the four pragmatic types, it is a demonstrative.

Note that the order in which the referent types are presented in Table 4.1 also reflects the diachronic development of demonstratives into definite articles as proposed by [Himmelmann \(1997\)](#). He argues against the common assumption that demonstratives develop into definite articles via anaphoric contexts, which are often treated as the bridge between deictic and uniqueness-based definite contexts (e.g. [Croft 2003: 246](#), [Greenberg 1978: 252](#), [Schroeder 2006: 554](#)). Instead, he views the anamnestic function, i.e. recognitional and via extension establishing, as the relevant one from which a demonstrative can develop into a marker of uniqueness-based definiteness, i.e. a definite article ([Himmelmann 1997: 93-97](#)). I follow [Himmelmann \(1997\)](#) in assuming that recognitional and establishing uses form the contexts from which demonstratives can (but do not have to) develop into definite articles. I therefore also assume that definite articles should generally be able to encode these two functions. Since a detailed diachronic discussion lies outside of the scope of this study, and since demonstratives are also able to occur in those two contexts, they will not play a role in defining definite articles for the purposes of the present study.

Another remark is necessary about situational uses shown in the first row of Table 4.1. It groups together deictic and situationally unique contexts because both involve a referent that is unambiguously identifiable in the immediate discourse situation. This goes back to [Hawkins \(1978\)](#), who at the same time also argues that demonstratives can only encode deictic referents, whereas situationally unique referents are expressed by the definite article ([Hawkins 1978: 111-115](#)). Thus, although the two types of contexts are very similar, this observation suggests that definite articles do in fact not occur with deictic referents in the strict sense, but mark the referent as situationally unique. The data from anaphoric and weak definite articles presented in Chapter 5 also supports the distinction of deictic and situationally unique contexts, since the contexts are not necessarily expressed in the same way by articles.

As was mentioned in the previous section, the distinction of the referential function of an expression in discourse is blurred by the fact that expressions can have or be compatible with more than one function at a time. Deictic referents and situationally unique referents are especially similar in their functions: the former draws the hearer's attention to a physical object in the discourse situation, and the latter marks the referent as the only salient referent of its kind in the immediate discourse situation. I will show in Section 5.2 that we find definite articles that are truly

incompatible with deictic referents. Therefore, it is very likely that definite articles occurring in situational uses as shown in Table 4.1 do in fact always mark the referent as situationally unique and not as deictic.

Combining these observations with the referential functions defined in Section 3.2, Table 4.2 shows the referential functions from the definite domain and their expression by demonstratives and (different types of) articles.

Table 4.2: The cut-off point between demonstratives and articles

marker	R_{deictic}	R_{anaph}	R_{recog}	R_{establ}	$R_{\text{rel-bridge}}$	$R_{\text{sit.unique}}$	$R_{\text{cont. unique}}$	$R_{\text{u-bridge}}$
DEM	✓	(✓)	(✓)	(✓)	(✓)			
ART:DEF	(✓)	✓	✓	✓	✓	✓	✓	✓
ART:ANA		✓	(✓)	(✓)	(✓)	(✓)		
ART:DEF _{WEAK}						✓	✓	✓
ART:RECOG			✓	(✓)				

Some demonstratives may be used in deictic contexts only, but often, demonstratives are also used to mark referents as anaphoric, and by extension, as recognitional and establishing. While a more thorough investigation would go beyond the purposes of the present study, I assume that demonstratives may also be able to encode at least certain types of rel-bridging referents (whether or not they do is not relevant for the purposes of this section).² The brackets in Table 4.2 indicate that this does not need to hold for each demonstrative within a given language; languages often have more than one demonstrative (e.g. with a distance contrast) and it may be the case that only one of them is extended to be used in recognitional, establishing, and rel-bridging contexts. In fact, [Himmelman \(1997: 71\)](#) notes that if a language has a proximal and a distal demonstrative, it is often (although not necessarily) the distal one that can be found in recognitional and establishing contexts.

For definite articles, Table 4.2 essentially follows the pattern shown in Table 4.1. One additional distinction in Table 4.2 is the one between deictic and situationally unique referents. Definite articles are marked as optionally compatible with deictic referents. They may occur in contexts that are ambiguous between situationally unique and deictic referents. Moreover, many languages use definite articles together with a demonstrative marker, but we also find languages in which definite articles are incompatible with deictic referents (cf. the examples shown in Sec-

²For instance, *this/that* are acceptable in the following example, adapted from [Schwarz \(2009: 279\)](#), which includes a rel-bridging context: *I travelled around South America last year. I want to tell you about my experiences on **this/that** trip.* On the other hand, examples like *John bought a book. #**This/that** author is French* are clearly much less acceptable, probably also due to the close proximity between the antecedent and the bridging referent.

tion 5.1.3). Other than deictic referents, definite articles are assumed to be able to encode all referent types listed in Table 4.2, on the basis of the definition of definite articles given in Chapter 3.

Another type of article distinguished here is the anaphoric article. These markers have been discussed as demonstratives rather than as articles in the (typological) literature (e.g. Diessel 1999, Dryer 2013b, 2014, Himmelmann 1997, De Mulder & Carlier 2011). For the purposes of the present study, I argue that anaphoric markers which do not mark referents as deictic differ from demonstratives. Whether or not one wants to argue that they represent a more abstract, extended, or grammaticalized function, adnominal anaphoric markers are clearly less basic in the sense that they are not found in all languages, whereas deictic demonstratives are. Often, it is the latter that can extend its referential functions to mark referents as anaphoric. Therefore, I treat markers that encode anaphoric referents but not deictic ones as anaphoric articles here, which is not very different from Schwarz (2009) who calls them as strong definite articles. Also Himmelmann (1997: 66) notes that treating the anaphoric marker in Wubuy (Nunggubuyu, Australia) as a demonstrative might be problematic, since it cannot be used to mark spatial deixis and thus differs considerably from the functions usually ascribed to demonstratives. In addition, he also mentions that such markers may rather be classified as anaphoric articles (Himmelmann 1997: 194), which is exactly what I propose here.

Anaphoric articles are mainly used to encode anaphoric referents. On the one hand, they cannot be used to mark deictic referents which distinguishes them from demonstratives. On the other hand, they do not express uniqueness-based functions in the definite domain on the right end in Table 4.2, i.e. they do not mark referents as contextually unique or u-bridging. Usually, anaphoric articles are not to express situational uniqueness either, but as will be shown, some anaphoric articles can do so. These contexts distinguish them from definite articles.

The occurrence of anaphoric articles with recognitional referents and establishing referents is indicated in brackets in Table 4.2. Expressing these two functions may not be part of the main function of an anaphoric article; however, a number of anaphoric articles from the sample can in fact be used to encode recognitional and establishing referents (cf. Section 5.2). Moreover, as was mentioned above, Himmelmann (1997) argues that these contexts are the ones through which definite articles develop. In other words, we can view them as the intermediate functions between pragmatic and semantic definites. It is therefore not surprising that anaphoric articles, which are used to express pragmatic definite functions, may in some cases extend their use towards the coding of recognitional and establishing (and situationally unique) referents.

Weak definite articles, also shown in Table 4.2, are complementary to demonstratives and anaphoric articles in the sense that they only occur with semantic definite referents. The last type of articles in the definite domain is the recognitional article. The reasoning for classifying markers

that are used exclusively to express a recognitional function is the same as for anaphoric articles. Importantly, recognitional articles do not encode deictic or anaphoric referents (in contrast to demonstratives or anaphoric articles), and they do not occur in uniqueness-based definite contexts either (in contrast to definite articles). Recognitional articles can however also be used to establish a referent in the discourse (cf. Section 3.4). Given that the sample only contains five recognitional articles, it is difficult to say whether or not this is necessarily the case; therefore, the establishing use of recognitional articles is indicated in brackets in Table 4.2.

As was mentioned above, the contexts of deictic and situationally unique referents overlap and cannot (or may not need to) be disambiguated in most cases. For practical reasons, I will therefore not consider situationally unique referents as a criterion to distinguish between demonstratives and definite articles in the remainder of this study. Two other referential functions that are not fully included in the set of defining contexts of definite, anaphoric, and weak definite articles are establishing and recognitional contexts. Establishing contexts are assumed to be expressed by demonstratives, definite, anaphoric, and recognitional articles, and are therefore not a useful diagnostic tool to distinguish between these markers. The same holds for recognitional referents as one of the defining contexts for definite articles. As shown in Table 4.2, definite articles are generally expected to be able to encode a recognitional function. However, such uses are often not provided in language descriptions, so that I exclude recognitional referents here for practical reasons as well. This is also true for bridging functions; these are defining for definite articles but are not always found in the language sources, which is why I do not include them as a necessary criterion here either.

This leaves us with the referential functions shown in Table 4.3 that are minimally required to distinguish between definite, anaphoric, weak definite, and recognitional articles. These are also the criteria that were applied to all articles from the definite domain in the sample. Again, brackets indicate that the use of the marker in a given context can vary. The contexts shown in gray are those which are irrelevant for the distinction between article types and demonstratives as explained in the previous paragraphs.

Table 4.3: Defining criteria for the articles in the definite domain

marker	R_{deictic}	R_{anaph}	R_{recog}	$R_{\text{cont.unique}}$
demonstrative	✓	(✓)	(✓)	✗
anaphoric article	✗	✓	(✓)	✗
weak definite article	✗	✗	✗	✓
recognitional article	✗	✗	✓	✗
definite article	(✓)	✓	✓	✓

Demonstratives are expected to mark deictic referents and they are expected to be infelicitous to encode contextually unique referents. What I treat as an anaphoric article occurs with anaphoric referents but at the same time cannot mark deictic or contextually unique referents. For recognitional articles, the important criteria are that they occur with recognitional referents but not with deictic, anaphoric, and contextually unique ones. Markers that systematically encode anaphoric and contextually unique referents are treated as definite articles. The following paragraphs will briefly present each of these markers.

In order to count as a demonstrative, the marker has to occur systematically with deictic referents and it needs to be infelicitous with contextually unique ones. Although this type of marker is not relevant for the remainder of the present study, an example will be provided here to illustrate how the criteria shown in Table 4.3 were applied to exclude markers. Ejagham (Bantu, Cameroon) has a demonstrative that is used for deixis (137) and anaphora (138).

- (137) ma ŋɔle gibe ka [ečo ndjik] elogo.
 1SG beg hold for.me bag.CL5 DEM:PROX.CL5 small
 ‘Please, can you hold this bag for a second?’
 Ejagham (primary data)

- (138) a. ewugu m-ba ŋgi mɔninki na aβɔ aβi ka kifən.
 when 1SG-came saw woman.CL1 with 2 children in kitchen
 ‘When I came home, I saw a woman with two children in the kitchen.’
 b. bo mɔpɛ wi [mɔninki ŋu]?
 PST know.2SG CL1 woman.CL1 DEM:PROX.CL1
 ‘Did you know this woman?’
 Ejagham (primary data)

We can argue that this marker in Ejagham is a demonstrative and not a definite article because it cannot mark referents as contextually unique. Example (139) shows this for the referent of *ndzuk oβasi* ‘church’ which is unique given the larger context of the village the discourse situation is placed in.

- (139) [ndzuk oβasi (*ŋi)] erifa ka etek ndjik?
 house.CL9 god DEM:PROX.CL9 where in village.CL5 DEM:PROX.CL5
 ‘Where is the church in this village?’ Ejagham (primary data)

In order to count as an anaphoric article, a marker has to occur with anaphoric referents but not with deictic or contextually unique referents. An example of such a marker can be found in Limbum, another Bantu language spoken in Cameroon. In (140), we see that the marker *fɔ̃* is used in anaphoric contexts, namely when the former hearer refers back to *ndjínwě fɔ̃* ‘that woman’ in (140b).

- (140) a. mú mfè? mē mū bā? mē mū yé ndjɪŋwè bá bōō bāā ó ćútèh mí
 when time 1SG PST2 arrive 1SG PST2 see woman and child.PL two 3PL sit in
 kíŋin.
 kitchen
 ‘When I came home, I saw a woman with two children in the kitchen.’
- b. wè mū rīŋ [ndjɪŋwě f̄ɔ / *ćà] ā?
 2SG PST2 know woman ART:ANA / DEM:DIST Q
 ‘Did you know that woman?’
 Limbum (primary data)

On the other hand, examples (141) and (142) below show that *f̄ɔ* cannot be used with deictic or contextually unique referents. It is therefore treated as an anaphoric article here (cf. Section 5.2.1 for more details). In the deictic context in (141), the distal demonstrative *ćà* needs to be used instead. With contextually unique referents such as *ntāh* ‘market’ in (142), *f̄ɔ* cannot be used; the referent needs to be expressed by a bare noun instead.

- (141) wè rīŋ [ndjɪŋwě *f̄ɔ / ćà] ā?
 2SG know woman ART:ANA / DEM:DIST Q
 ‘Do you know that woman (over there)?’
 Limbum (primary data)
- (142) [ntāh (*f̄ɔ)] yō fē à là? nā?
 market.CL9 ART:ANA be.CL9 where in village this?
 ‘Where is the market in this village?’
 Limbum (primary data)

Recognitional articles encode only recognitional referents, as *max* in Oksapmin (Oksapmin, Papua New Guinea). The language has an elaborate system of various clitic and free adnominal spatial demonstrative markers, besides a definite, an indefinite, and a recognitional article (Loughnane 2009: 104-130). An example of the recognitional article is given in (143):

- (143) gin i ml-sa jəxe tumbuna paxna sup [stori max] pla gina
 now HESIT come.up-SEQ then ancestor hunger illness story ART:RECOG tell.PRS.SG now
 ...
 ...
 ‘Now, I came up and told that story about famine in the old days. Now ...’
 Oksapmin (Loughnane 2009: 124)

Here, the speaker refers to *story max* ‘that story’ that they had told the hearer in a different conversation. The expression used in (143) is the first mention of the story in the current discourse situation. Example (144) shows that the language uses a different demonstrative, proximal *mə=* in this case, to express deictic referents.

- (144) ox mǝ=tǝx pa se gno-m pt-sux=li
 3SG.M DEM:PROX=place taro INFER grow-SEQ be-HAB.EVID.REMPST=REPORT
 ‘I guess he must have stayed here [this place] growing taro.’
 Oksapmin (Loughnane 2009: 110)

The remaining two relevant referent types, anaphoric and contextually unique referents are coded by the definite article *jox*. Examples (145) and (146) illustrate this, respectively. Of course, such examples as well as example (144) do not exactly provide evidence for *max* being infelicitous in these contexts, but they reflect its general absence throughout the examples provided in Loughnane (2009). Moreover, Loughnane (2009: 123-125) discusses the marker as a “dedicated recognitional demonstrative” (cf. Section 5.4.1 for more details).

- (145) a. [ap tit] x-t x-n-gop=li ej
 house ART:INDEF do-PFV-EVID.HODPST.SG be-PFV-VIS.REMPST.SG=REPORT gosh!
 x-ti-p x-n-gop=li je put te put
 do-PFV.EVID.REMPST.SG be-PFV-VIS.REMPST.SG=REPORT mountain top place top
 te=nǝp mox ...
 place=INT DEM:PROX ...
 ‘There was a house which had just been built, sorry, which had been built long ago.
 Right at the very top of that mountain ...’
- b. ...ux [ap jox] loj-xi-p=li
 3SG.F house ART:DEF enter-PFV-EVID.REMPST.SG=REPORT
 ‘She went into the house.’
 Oksapmin (Loughnane 2009: 129)

- (146) tom jox lum p-d-m edi-pla=o [ake jox]
 water ART:DEF a.lot CAUS-eat-SEQ stay.PFV-REMFUT.SG=QUOT stomach ART:DEF
 ox=o tom=wi x-ti-plox=xejox n-pli-nuŋ.
 2SG.M=QUOT water=only be-PFV.HODFUT.SG=because O:2-tell-EVID.HODPST.SG
 ‘Don’t give her too much water! Her stomach will fill up with water, she told me.’
 Oksapmin (Loughnane 2009: 480)

For definite articles, I argued that the defining functions are their occurrence with anaphoric and contextually unique referents. The last two examples from Oksapmin, (145) and (146), show this for *jox*. In (357a), the speaker first introduces the referent of *ap* ‘house’ using the indefinite article *tit*. When mentioning ‘the house’ again in (357b), the speaker refers to ‘the house’ using the definite article *jox*. In example (146), we see that the same marker, *jox*, is used to mark the referent of *ake* ‘stomach’ as contextually unique. The referent is contextually unique based on world knowledge that a person has one and not more than one stomach.

To conclude, this section introduced the criteria that are necessary in order to distinguish definite, anaphoric, and recognitional articles from demonstratives³. The last part gave examples of how the criteria can be applied to markers from different languages; this was applied to all the articles in the sample that belong to the definite domain. While only a selection of articles in the definite domain are presented in Chapter 2, examples of the relevant contexts of all articles in the sample can be found in the online appendix.

4.2 Possessives and definite articles

Adnominal possessives correspond to another type of markers that have been argued to be a source of definite markers or definite articles due to their functional overlap with definite articles in other languages (e.g. Krámský 1972 for Uralic, Turkic, Armenian, Javanese). This relation between possessive and definite markers was discussed in more detail in the literature especially for Uralic (e.g. Fraurud 2001, Gerland 2014, Nikolaeva 2003, Serdobolskaya et al. 2019) and Semitic (e.g. Appleyard 2005, Huehnergard & Pat-El 2012, Pat-El 2009, Rubin 2010), as well as for Yucatec Maya (C. Lehmann 2003), Turkish (Schroeder 1999), and Indonesian (Gesita & Suhardijanto 2020, Rubin 2010). In this section, I will focus on data from Udmurt (Uralic), Amharic (Semitic), and Indonesian (Austronesian) to discuss the similarities and differences as well as the diachronic relation between definite articles and the possessive markers. Following Fraurud (2001), Gerland (2014), Nikolaeva (2003), Serdobolskaya et al. (2019), I will argue that even though possessive markers can occur in all contexts in which a definite article is expected to occur, they cannot count as such because they do not seem to be used systematically in such contexts and are often described as purely optional. This argument made for Udmurt and Indonesian generalizes to the other Uralic languages and Turkish as well. The situation in Semitic, e.g. Amharic, is somewhat different. Amharic has a definite article which was assumed to have developed from a possessive marker; at the end of this section I will briefly summarize the argument made in Huehnergard & Pat-El (2012) against such a development. I also show that a similar argument can be applied to marker in Indonesian as well.

4.2.1 Udmurt

Following Aikhenvald & Dixon (2013: 3), the core domain of possessive markers or possession includes ownership (of property), whole-part relations (including body parts, plant parts), and kinship relations. In addition, in some languages possessive markers or constructions are used to mark associations in a more general sense. Aikhenvald & Dixon (2013: 3) provide the following

³Weak definite articles are in complementary distribution with demonstratives, which is why they were not shown here. Section 5.3 discusses weak definite articles in detail.

example: an expression such as *John's house* could refer to *the house John owns*, or *the house he lives in (but does not own)*, or *the house he is associated with (for example, designed, or built, or drew a picture of)*. It is especially this latter use to express a general association between two referents that shows functional overlap with definite articles marking bridging and contextually unique referents. Hence, if one wants to posit a development from a possessive to a definite marker, one may expect that the possessive marker, expressing general associations in bridging and contextually unique contexts first, becomes a systematic marker in such contexts. In order to develop into a definite article, its use would then also need to extend to the non-uniqueness-based (or pragmatic) definite contexts, i.e. one would expect to find the marker with establishing, anaphoric, and situationally unique referents. Once it is used systematically in such contexts as well, it could be treated as a definite article.

At a first glance, the possessive markers in various Uralic languages seem to be undergoing this development. Previous studies have shown that possessive markers occur in many contexts that definite articles in other languages occur in (Fraurud 2001, Gerland 2014, Nikolaeva 2003, Serdobolskaya et al. 2019, and references therein). Gerland (2014) lists the following Uralic languages for which this was observed: the Ob-Ugric languages Mansi and Khanty; the Permic languages Udmurt, Komi, and Mari; and the Samoyedic languages Nenets, Enets, Nganasan, and Selkup (Gerland 2014: 271).

Nikolaeva (2003) and Gerland (2014) provide a detailed overview of the use of possessive markers in several of these languages in different definite contexts. Their data shows that possessive markers from different Uralic languages do indeed occur in the contexts that were listed as contexts in which definite articles systematically occur in Section 4.1. A similar distribution of possessive markers in the Beserman variety of Udmurt is shown by Serdobolskaya et al. (2019). Thus, in a number of Uralic languages, possessive markers are used with anaphoric, establishing, situationally unique, contextually unique, and bridging referents. Therefore, it is tempting to analyse such possessive markers as (emerging) definite articles.

As was mentioned in the beginning of this section, Nikolaeva (2003), Gerland (2014), and Serdobolskaya et al. (2019) all show and conclude that the use of the possessive markers is by no means obligatory or systematic in such contexts. It is rather blurred by a number of additional pragmatic and other semantic factors. I follow the argumentation of the authors in that I only treat markers as definite articles if they systematically occur with (at least) anaphoric and contextually unique referents. In other words, if such referents are expressed by nouns or similar lexical referential expressions, the definite article has to be used as well; its absence in such contexts should be infelicitous. This does not seem to be the case for the Uralic languages discussed in Fraurud (2001), Gerland (2014), Nikolaeva (2003), Serdobolskaya et al. (2019). Thus, the crucial difference between such possessives and definite articles is neither the use of possessives in

possessive contexts nor the additional pragmatic functions the possessives have. The important difference between those possessives and definite articles is that the possessives do not mark the definite referent types in a systematic way. The following paragraphs will show this for Udmurt.

In Uralic languages, adnominal possessive markers are suffixed to the noun expressing the possessum. This is shown for Udmurt in (147).

- (147) *nə̃l-ə̃z*
 girl-**POSS:3SG**
 ‘his/her daughter’
 Udmurt (Serdobolskaya et al. 2019: 291)

Udmurt uses different markers with direct objects (DO) vs. other syntactic contexts (non-DO), and it marks alienable and inalienable possession in non-DO contexts. Because this section will mainly focus on third person possessive markers, the relevant third person markers are given in Table 4.4 following Serdobolskaya et al. (2019: 295).

Table 4.4: Third person possessive markers in (Beserman) Udmurt

context	possessive marker
DO	<i>-ze / -se</i>
non-DO inalienable	<i>-(j)ə̃z / -z</i>
non-DO alienable	<i>(j)ez</i>

Given that possessive markers are inherently able to express (abstract) associations between referents, the functionally closest types of definite contexts are contextually unique and bridging contexts. That the Udmurt possessive can mark contextually unique referents is shown in (148). Here, *-se* is used to mark a contextually unique referent, namely the referent of *kompl’eks* ‘[cattle breeding] complex’ in Shamardan, which is the village where the discourse situation takes place, functioning as its larger context.⁴

- (148) *ten’ ta-tə̃n as-la-m Šamardan-ə̃n kompl’eks-se vuza-zə̃, tak*
 here this-LOC REFL-GEN1-P.1 Shamardan-LOC complex-**POSS:3SG.ACC** sell-3PL.PST so
 ved’?
 PTCL
 ‘[Yes, Mikhalych, why speaking of Moscow and Leningrad,] when even here in Shamardan
 they [the officials] have sold the [cattle-breeding] complex, haven’t they?’
 Beserman Udmurt (Serdobolskaya et al. 2019: 302)

⁴Serdobolskaya et al. (2019) mention that *kompl’eks* ‘complex’ is introduced as the new discourse topic in (148) and that new discourse topics tend to receive possessive marking in Beserman Udmurt.

The utterances in (149) and (150) show how possessive markers are used in bridging contexts to mark the relation between a previously mentioned referent and the bridging one. Example (149) shows a description of preparing pickled cabbage. The referent of *rassol* ‘brine’ is an identifiable bridging referent because of its link to the previously mentioned pickled cabbage. In (150), the speaker first mentions the event of mushroom picking, serving as the anchor of the bridging referent of *gibi* ‘mushroom’, marked by *-ze*. According to Serdobolskaya et al. (2019: 300), the DO-set (i.e. the accusative form) of possessive markers is systematically used in such bridging contexts.

- (149) Sre rassol-ze kiš't-iš'ko.
 then brine-POSS:ACC.3SG pour-PRS.1SG
 ‘[I make pickled cabbage. First I immerse it in water, ...add salt, crush it ...]
 Then I pour out the brine.’
 Beserman Udmurt (Serdobolskaya et al. 2019: 300)

- (150) Tolon vel't-i-m č'aš'a-e gibija-nâ, so bere č'aš'a
 yesterday go-PST-1PL forest-ILL pick.mushrooms-INF that after forest
 gibi-ze ujbât š'už'ja-m.
 mushroom-POSS:ACC.3SG whole.night clean-PST.1PL
 ‘We went to the forest to pick mushrooms and then we spent the whole night cleaning
 [the] mushrooms.’
 Beserman Udmurt (Serdobolskaya et al. 2019: 306)

Assuming the development from possessive to definite marker, one would expect the marker in question to occur with pragmatically definite referents such as anaphoric referents as well. This is indeed the case. Example (151) shows the possessive *-jez* marking an anaphoric referent as identifiable.

- (151) Žič'i-jez tare košk-i-z ot-âš'.
 fox-POSS:3SG then go.away-PST-3SG there-EL
 ‘[Then take a picture of the fox and put it near the girl with yellow hair.]
 Then the fox went away from there’
 Beserman Udmurt (Serdobolskaya et al. 2019: 300)

Although anaphoric referents can be expressed by a lexical noun with the possessive marker in Udmurt, the latter is not required to be used in such contexts (Serdobolskaya et al. 2019). Consider the anaphoric referent of *gondâr* ‘bear’ in (152b), which does not show any additional marking and is expressed by a bare noun.

- (152) a. Odig pol mân-e až'-e gondâr-ez.
 one time go-PRS.3PL see-PRS.3SG bear-ACC
 ‘[The fox is driving a cart.] Once, while driving, it sees a bear.’

- b. **Gondâr vera:** “*Ž’ič’â puk-t-â mon-e!*”
bear tell.PRS.3SG fox sit-CAUS-IMP.SG I-ACC
 ‘The bear says: “Fox, give me a lift!”’
 Beserman Udmurt (Serdobolskaya et al. 2019: 307)

Moreover, Serdobolskaya et al. (2019: 307) mention that their corpus contains 12 contexts similar to the one in (152), and that none of them features the possessive marker on the noun expressing the anaphoric referent. This suffices to conclude that the possessive marker in Beserman Udmurt is not a definite article. Nikolaeva (2003) provides similar examples for Komi, Selkup, and Khanty, showing that the possessive marker can but does not have to be used with anaphoric referents.

Nevertheless, note that the Udmurt possessive can occur with deictic or situationally unique referents, as is shown in (153). Here, the possessive is used with a noun that is also marked by the distal demonstrative *so*, so we can conclude that the possessive is at least compatible with a deictic or situationally unique interpretation of the referent.⁵

- (153) **Kin’ dor-e? – So lud ad’ami-ed dor-e.**
 who near-ILL **that** wild man-**POSS:2SG** near-ILL
 ‘To whom? – To that wild man.’
 Beserman Udmurt (Serdobolskaya et al. 2019: 304)

However, Serdobolskaya et al. (2019: 307) mention that the use of the possessive is by no means systematic or obligatory with deictic or situationally unique referents. They provide example (154), showing another situationally unique referent marked by a demonstrative, this time without the possessive marker.

- (154) **Ta d’erevn’a-jân mon vorž’-ik-i.**
this village-LOC I.NOM give.birth-DETR-PST.1SG
 ‘I was born in this village.’
 Beserman Udmurt (Serdobolskaya et al. 2019: 307)

To sum up, even though the possessive marker in Udmurt can occur in all contexts relevant for the definite article, its use in such contexts is not systematic. In general, the distribution of the possessive marker in Udmurt rather follows other pragmatic functions. In addition to marking relations between referents, Udmurt possessives can introduce new discourse topics, mark contrastive topics, and emphasize deixis expressed by demonstratives according to Serdobolskaya et

⁵The possessive in (153) is a second person possessive marker. For the purposes of the present study I do not discuss in more detail the use of first and second person possessive markers, which can mark referents as definite similarly to third person possessives in the examples shown in this section. Using first or second possessives additionally associates the referent with the speaker or the hearer, also in a more abstract way than in a possessive relation. For more details on such uses of first and second person possessives in Uralic, see Nikolaeva (2003) and Serdobolskaya et al. (2019).

al. (2019). Moreover, first person possessives mark vocatives and second person possessives can be used to express the relation between the hearer and the utterance.⁶ Together with its inherent function to mark (abstract) relations between referents, all these factors lead to a distribution that overlaps with the one of definite articles in other languages.

Gerland (2014: 283) explains their use as a marker of definiteness by analysing the Uralic possessives as relational markers, whose main function it is to establish some kind of relation between two entities. A similar point is made in Nikolaeva (2003) and Fraurud (2001), who contend that the occurrence of the possessive markers in various types of definite contexts goes back to the possessive meaning of the markers, which is much wider and more abstract than what we find in Indo-European languages. In the latter case, the referent of the possessor often has to be established so that the possessive relation can be marked. Fraurud (2001), Gerland (2014), Nikolaeva (2003) argue that this is not the case in Uralic, which allows the markers to express a wider variety of relations between referents. Thus, the fact that we see certain possessives in similar contexts as definite articles can be taken to be a by-product of their semantic and pragmatic properties rather than the systematic expression of referential functions per se.

4.2.2 Amharic

As was mentioned in the beginning of this section, the case of Semitic is somewhat different. I will only provide a very brief summary of one important point here; see Appleyard (2005), Huehnergard & Pat-El (2012), Pat-El (2009), Rubin (2005, 2010) for diachronic and synchronic details on the definite article in different Semitic languages.

Semitic, especially Amharic, may be considered a showcase for possessive markers that have developed into definite articles. This is because Amharic has a definite article whose form is identical to the adnominal third person possessive marker. This observation lends itself to assume a diachronic relation between the two, i.e. a development from the possessive to the definite article. However, Huehnergard & Pat-El (2012) point out that there is a crucial difference in the syntactic behaviour of the two markers. Consider *bet-u* in (155a), which can either be interpreted as ‘his house’ or ‘the house’. Amharic nouns inflect for gender, which requires agreement with the definite article. The examples in (155a) and (155b) only allow for this ambiguity between a possessive and a definite interpretation because the possessor and the head noun are of the same gender. If the marker of the possessor and the head noun do not match in their gender value, only a possessive interpretation is possible. Thus, the definite article in Amharic agrees with the head

⁶Serdobolskaya et al. (2019) call the latter function the “ethical function” of possessives in Udmurt. It is also discussed in Nikolaeva (2003), and Gerland (2014: 277) notes: “Therefore, it seems to be the most plausible solution to assume that the application of the 2nd person establishes a closer link between hearer and the marked referent. Similar uses of the 2nd person possessive suffix can be found in many languages.”

noun, while the possessive marker, otherwise identical in form, marks the gender of the possessor which can differ from that of the noun it occurs with.

- (155) a. *bet-u* ‘his house (m.)’ or ‘the house (m.)’
b. *gäräd-wa* ‘her maid (f.)’ or ‘the maid (f.)’
c. *bet-wa* ‘her house (m.)’
d. *gäräd-u* ‘his maid (f.)’
(Huehnergard & Pat-El 2012: 39)

Huehnergard & Pat-El (2012) note that this different syntactic behavior is better explained by assuming two separate developments leading to the possessive marker and the definite article present in Modern Amharic today. They provide additional historical evidence and evidence from other Semitic languages supporting this analysis, reconstructing a demonstrative marker as the common source for both the definite article and the possessive marker. First, this automatically accounts for their different syntactic behavior that would have required an additional explanation under the assumption that the article developed from the possessive. Second, it also suggests that the definite article in Amharic (and other Semitic languages) does in fact not originate from a possessive marker but is only yet another example of the well attested grammaticalization path from a demonstrative to a definite article. This kind of dual development from a demonstrative to both a definite article and a possessive may also explain that the latter two markers have identical forms, which probably also applies to similar patterns in languages outside of the Semitic family.

4.2.3 Indonesian

Indonesian is another language in which the possessive marker *-nya* may be an emerging definite article or a marker of definiteness and identifiability of some sort (Englebretson 2003, Gesita & Suhardijanto 2020, Rubin 2010). The use of *-nya* in Indonesian (especially rather colloquial varieties) is comparable to the use of possessives with definite referents in Udmurt discussed in Section 4.2.1 in that it is governed more by discourse-pragmatic structures and is not required to mark a referent as identifiable. However, *-nya* also has similarities to the definite article in Amharic; it is not only a possessive marker but has many other functions. This section will focus on this latter property, leaving aside the question of how systematic the use of *-nya* with definite referents is, and showing that Indonesian *-nya* does not provide strong evidence for the development of a possessive marker in the strict sense into a marker of definiteness either.

The use of *-nya* as a possessive marker is shown in (156) and (157). Even though *-nya* has a number of other functions, it is systematically glossed as a possessive in this section for the sake of simplicity.

(156) Dia mengganjal-kan bantal itu ke punggung-nya.
 3SG wedge-APPL pillow DEM to back-POSS
 ‘He wedged that pillow under his back.’
 Indonesian (Englebretson 2003: 33)

(157) Nama-nya kan Rifka.
 name-POSS PRT Rifka
 ‘Her name was Rifka.’
 Indonesian (Englebretson 2003: 158)

Examples (158) and (159) show *-nya* in contexts where it does not mark a possessive relation. In (158), it is used with a rel-bridging referent, and we see in (159) that it can mark a referent as anaphoric as well.

(158) a. E=kalau mendoa-kan orang yang beda agama, boleh nggak.
 um=if pray-APPL person REL different religion may NEG
 ‘As for praying for someone of a different religion, is that okay or not?’

b. Tergantung doa-nya.
 depend prayer-POSS
 ‘Depends on the prayer.’
 Indonesian (Englebretson 2003: 165)

(159) a. jadi gua ntar ketemu dia langsung di salon.
 so I soon meet her direct in salon
 ‘So I’m going to meet her directly at the salon.’

b. salon-nya di dekat rumah?
 salon-POSS in near house
 ‘Is the salon near your house?’
 Indonesian (Rubin 2010: 109)

Besides its use as a marker of possession and as a marker of identifiability, Englebretson (2003: 153-186) provides a detailed description of the other syntactic and information structural functions of *-nya*. For instance, *-nya* also functions as a nominalizer, as is shown in (160).

(160) Mereka turun-nya di=itu kali Code situ.
 3PL get.off-POSS at=DEM river Code there
 ‘They got off the bus there at the Code river.’ (lit. ‘Their getting off was at the river Code there.’)
 Indonesian (Englebretson 2003: 168)

Without going into the morphosyntactic details here, *-nya* can also occur on certain types of verbs to cross-reference a third person patient. This is shown in (161), (162), and (163).

- (161) Saya menganggap-*nya* bodoh.
 1SG consider-**POSS** stupid
 ‘I consider him stupid.’
 Indonesian (Sneddon et al. 2010: 169)
- (162) Dia mengambil kue itu lalu memakan-*nya*.
 3SG take cake DEM then eat-**POSS**
 ‘He took the cake and ate it.’
 Indonesian (Sneddon et al. 2010: 170)
- (163) Mati-in-*nya* gimana nih.
 die-APPL-**POSS** how this
 ‘How do you turn it off?’
 Indonesian (Englebretson 2003: 169)

Other uses of *-nya* include framing or backgrounding a situation (Englebretson 2003: 172-185), and it is also found in some types of adverbial expressions and equative constructions (Sneddon et al. 2010: 217-218, 186-187). Leaving aside the diachronic relation between these different functions, the aim of this section was to show that *-nya* is not only used as a possessive marker in the strict sense but should probably be analysed as a much broader nominal or pronominal marker. Therefore, even if *-nya* develops into a definite article (which itself is questionable), it is difficult to argue that this would be an example of a definite article originating from a possessive marker.

To sum up, the Udmurt examples suggested that even though possessive markers can show a distribution similar to the one of definite articles in other languages, they are not necessarily evidence for emerging articles that develop from possessives. The Amharic and Indonesian data provided further evidence against a clear development from a possessive to a definite article. Taken together, the crosslinguistic evidence for a development from possessives to definite articles is generally rather weak and needs to be revisited in future studies to provide a more conclusive picture.

4.3 The numeral ‘one’ and articles in the indefinite domain

The distinction between indefinite or exclusive-specific articles and the numeral ‘one’ is difficult to make in certain cases because exclusive-specific articles and indefinite articles almost always develop from the numeral ‘one’. To give an example, Romero-Méndez (2008: 270) notes the following for Ayutla Mixe (Mixe-Zoque, Mexico): “The numeral *tu’uk* ‘one’ is commonly used as an indefinite article.” Two examples to illustrate its use are given in (164) and (165), in both of which a not yet identifiable but discourse-prominent referent is introduced.

- (164) jam [tu'uk mäj tsäj].
 there **one** big stone
 'There was a rock.'
 Ayutla Mixe (Romero-Méndez 2008: 271)
- (165) kuu=äk=ja'a jä'äy y-men-y [te'n tu'uk tixytyëjk=äk
 when=QUOT=DEM person 3SG-come-DEP DEM **one** woman=QUOT
 poop=te'kn-äk=ja'a] y-men-y.
 white=INTENS=QUOT=DEM 3SG-come-DEP
 'When the people came, a woman in white came.'
 Ayutla Mixe (Romero-Méndez 2008: 271)

As the glosses indicate, I do not treat the marker as an article. As I show later in this section, I view this as the presentative use of the numeral 'one', which seems to be generally available in the world's languages if the language does not have a separate specific or indefinite article.

The gradual diachronic development from the numeral to an exclusive-specific or an indefinite article (cf. Heine & Kuteva 2002: 219-221; Himmelmann 2001; Dryer 2013b) is shown in Figure 4.2 following Heine (1997: 71-76).

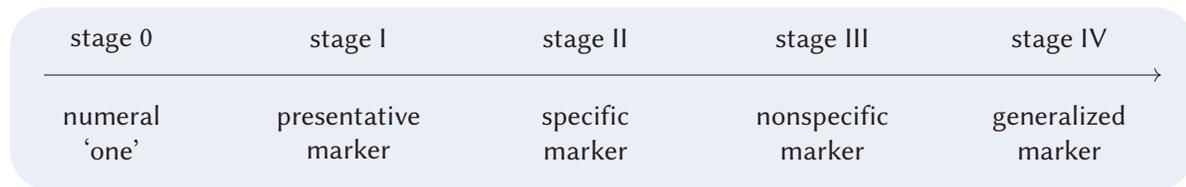


Figure 4.2: Grammaticalization path from the numeral 'one' to a nominal marker

The stages from the scale in Figure 4.2 that are relevant to the discussion of indefinite articles and the numeral 'one' are the stages 0 to III. Stage I markers, called "presentative markers" in Heine (1997), are markers that signal prominent, topical discourse referents as not (yet) identifiable to the hearer. Most instances correspond to the introduction of new discourse referents. Specific markers of stage II correspond to exclusive-specific articles. Nonspecific stage III markers correspond to indefinite articles, being used to encode both specific and nonspecific referents. The scale is usually assumed to be implicational if interpreted synchronically:

The model can be interpreted in a dual way. On the one hand, it may be viewed as a synchronic implicational scale. This means, for example, that an indefinite article of a given stage also has, or may have, the properties of all preceding stages, but not vice versa. On the other hand, it is claimed to reflect diachronic evolution, where the initial stage represents the oldest and the final stage the most recent situation. (Heine 1997: 71)

Instead of using the scale from Figure 4.2 as the basis, most definitions of indefinite articles that aim at a distinction between indefinite articles and the numeral ‘one’ consist of a list of typical features that contrast these two elements. To show an example without discussing it further, Table 4.5 contains the properties provided in Heine (1997) based on Heine et al. (1995). As such, the properties listed cannot be used to distinguish between the numeral ‘one’ and indefinite articles.

Often, distinct forms or stress patterns of the marker and the numeral are taken as sufficient to treat it as an article. Although the degree of form developments or phonetic erosion may reflect the degree of semantic changes and extensions, the form development is independent from the function development and cannot be a criterion as such. The necessary criteria should be based on the functions of the markers instead. Thus, the distinction between the numeral ‘one’ and the article can be made based on a cut-off point on the scale presented in Figure 4.2, using it as a synchronic hierarchy of related functions (and not necessarily as an indicator of a grammaticalization path). Only one stage from Figure 4.2 needs further clarification. As will be shown in Section 6.2, indefinite articles that are restricted to occur with discourse-prominent referents (called presentational articles) are common in the world’s languages. Their existence shows that the semantic development from specific to nonspecific contexts is independent from the pragmatic development of the article from more to less prominent referents. Therefore, an additional distinction needs to be made between prominent and less prominent contexts in the indefinite domain. The different functions distinguishing between the numeral ‘one’ and different types of articles in the indefinite domain are shown in Table 4.6. Because of the interaction between semantic and pragmatic distinctions, Table 4.6 can no longer be read as an implicational scale, but this is irrelevant for the purposes of this section.

Table 4.5: Characteristic properties of indefinite articles (Heine 1997: 68)

Length	Indefinite articles are generally short (≤ 2 syllables).
Stress	They are stressless.
Position	They are likely to employ the same position in the clause as the numeral ‘one’.
Noun types	They tend to be confined to determining the singular of count nouns. Nevertheless, there may be exceptions where the article has been extended to nonsingular referents.
mass > PL	If the indefinite article determines mass nouns, then it is also used for plural nouns.
PL > SG	If it determines plural nouns, then it also determines singular nouns.

Table 4.6: The cut-off point between the numeral ‘one’ and articles

	individuation	prominent R_{spec}	R_{spec}	prominent R_{nspec}	R_{nspec}
numeral ‘one’	✓	(✓)			
ART:EXSPEC	✓	✓	✓		
ART:INDEF	✓	✓	✓	✓	✓
ART:PRES	✓	✓		✓	
ART:NSPEC				✓	✓

The numeral ‘one’ has the main semantic function of individuation and quantification (cf. Givón 1981: 50), and it appears to be universal. Even though extremely simple numeral systems are attested, the simplest that are known still consist of a lexeme for ‘one’, sometimes along with words for ‘two’, ‘three’, or ‘few’ vs. ‘many’ (cf. Stampe 1976: 596; Greenberg 1978: 256; Dixon 1980: 107-108; Heine 1997: 24).⁷

For the presentative function of stage I as shown in Figure 4.2, i.e. the introduction of new (specific) referents into the discourse, we can observe a strong crosslinguistic trend towards using the numeral ‘one’ especially with single human referents (e.g. Givón 1980: 50; Heine 1997: 72; Dryer 2013b). Hence, this is a common property of the numeral ‘one’. There are two important restrictions that go with this use: first, the use of the numeral ‘one’ is confined to referents which are prominent and salient in the discourse; second, its use is not required or systematic but often emphatic. Therefore, the use of the numeral ‘one’ with prominent specific referents is marked in brackets in Table 4.6.⁸

An example to illustrate the presentational use of the numeral ‘one’ comes from Tikuna (Tikuna-Yuri, Colombia) in (166). Although the use of the numeral in such contexts is possible in Tikuna, it is not required in examples (167) and (168).

- (166) nücha dau [(wii) hipata hichi] lea wadiu-wa (naru) napaeru na
s:1SG.O:3SG see **one.PRES** house old DEM:DIST village-LOC POSS roof COP
chiee.
damaged
‘I see/saw a house in that village; its roof was damaged.’
Tikuna (primary data)

⁷The famous counter-example to this claim is Pirahã (Mura, Brazil). The two numeral markers *hói* and *hoi* were first analysed as lexemes for ‘one’ and ‘two’ in Everett (1992). This was supported by Gordon (2004), who claims that the language does not encode numerals above ‘two’. In a later study, Everett (2005) rejected the analysis of these markers as exact quantifiers, supported by results from a cognitive experimental study (Frank et al. 2008). Since it does not seem entirely clear what the most adequate analysis of the Pirahã numeral system is (cf. Everett 2009), I assume that the world’s languages generally have a lexeme that corresponds to the meaning of ‘one’.

⁸It is unclear at this point whether the numeral ‘one’ can also be used in nonspecific contexts. If so, its use would be assumed to be “optional”, i.e. non-systematic, as with specific referents.

- (167) i-pata-wa nucha dau [(wii) güena] tare buun pema cuchia-wa.
 1SG-house-LOC s:1SG.O:3SG see **one.PRES** woman two child there kitchen-LOC
 ‘When I came home, I saw a women with two children in the kitchen.’

Tikuna (primary data)

- (168) nukü dau [(wii) ail]!
 s:2SG.O:3SG see **one.PRES** tiger
 ‘Look! A tiger!’

Tikuna (primary data)

Because the numeral ‘one’ is widely attested in this function, I view the potential to occur with discourse-prominent specific referents as an extended function of the numeral ‘one’ that is generally available in the world’s languages, similarly to the anaphoric use of primarily deictic demonstratives. However, if new discourse-prominent referents are systematically encoded by a marker which may or may not be formally identical to the numeral ‘one’, and if this marker also occurs with discourse-prominent nonspecific referents, I treat it as a presentational article (cf. Section 6.2). Presentational articles therefore differ from the numeral ‘one’ because they can mark nonspecific referents, and because the systematic coding of specific or nonspecific discourse-prominent referents exceeds the non-systematic presentational occurrence that the numeral ‘one’ may have. Note that semantically, presentational articles are comparable to indefinite articles because they have the same referential functions.

Exclusive-specific articles are systematically used to mark a referent as specific, independently of its discourse prominence. Exclusive-specific articles can be found in e.g. Palula (Dardic, Pakistan). Examples (169) and (170) below show the article *áa(k)* with a discourse prominent referent; in (171), the noun *baṭ-á* ‘stone’ occurs with the exclusive-specific article although its referent is less prominent (cf. Section 6.3 for more examples).

- (169) bhuná [áa gíri] heensíl-i
 below **ART:EXSPEC** rock stay.PFV-F
 ‘Down below there was a big rock.’

Palula (Liljegren 2016: 309)

- (170) miír thaní [áak míiš] heensíl-u de.
 Mir **QUOT ART:EXSPEC** man stay.PFV-M.SG PST
 ‘There was a man called Mir.’

Palula (Liljegren 2016: 309)

- (171) eesé zangal-í [áa baṭ-á] ḡhulí harí so kuṇaák
 REM forest-OBL ART:EXSPEC stone-OBL on take.away.CV DEF.NOM.M.SG child.NOM.M.SG
 bheešóol-u.
 seat.PFV-M.SG
 ‘In that forest he took the child to a stone and seated him.’
 Palula (Liljegren 2016: 283)

Another type of articles listed in Table 4.6 is the indefinite article, which encodes specific and nonspecific referents, irrespective of their discourse prominence. The marking of both referents and the resulting ambiguity was shown for English in example (20) which is repeated here as (172). For more examples of indefinite articles, see Section 6.1.

- (172) John wanted to marry a rich woman ...
 a. ...but she refused him. (specific)
 b. ...but he couldn't find any. (nonspecific)

The last type of indefinite articles presented in Table 4.6 are nonspecific articles. These articles mark only nonspecific referents and do not occur with specific referents. An example of the nonspecific article in Ayoreo (Zamucoan, Bolivia) is shown in (173) below (cf. Section 6.4 on nonspecific articles).

- (173) Mu que ore i-plata-rigui cuse.
 but NEG 3PL 3-money-ART:NSPEC.M.SG EXIST
 ‘But they have no money.’
 Ayoreo (Ciucci 2016: 356)

As we saw in this section, the numeral ‘one’ and different types of articles from the indefinite domain can be distinguished based on the (referential) functions that the markers have, assuming the relevant functions to be individuation (quantification), and marking a referent as specific or as nonspecific. In addition, a pragmatic restriction along the lines of discourse prominence is important to distinguish presentational from indefinite articles.

4.4 Negative polarity items and nonspecific articles

Nonspecific articles are similar to negative polarity items (NPIs), and it may not be evident to distinguish these two markers. In contrast to the distinctions between articles and related markers discussed in the previous sections, there is no established diachronic relation between nonspecific articles and NPIs from the literature. Nevertheless, I will briefly describe the similarities and differences between those two types of markers, also to clarify what should and should not count as

a nonspecific article. Essentially, adnominal NPIs and nonspecific articles occur in the same contexts, but while the NPI usually comes with additional pragmatic functions and is not obligatory, nonspecific articles have to occur when nouns are used to express nonspecific referents.

The relevant types of NPIs are those that occur in the nominal domain together with a noun, as does *any* in English, e.g. in *Peter did not see any sign*. Since NPIs have been a popular topic in semantics for the last 50 years, I cannot mention and summarize all the aspects of NPIs that the literature has dealt with.⁹ The relevant uses of adnominal NPIs cover a wide range of functions (Haspelmath 1997: 52; Givón 1984: 441-449; Giannakidou 2017). Table 4.7 shows these functions; as becomes evident from the examples of nonspecific articles in Section 6.4, nonspecific articles occur in exactly the same contexts. Thus, what needs to be explained is when a marker that occurs in these contexts counts as an NPI and when it counts as a nonspecific article.

Table 4.7: Relevant contexts for adnominal NPIs

negation	direct	<i>I didn't see any cake in the fridge.</i>
	indirect	<i>I don't think there was any cake in the fridge.</i>
		<i>Tabea wrote a dissertation without having any coffee.</i>
	inherent negative predicates	<i>Tomas failed to write any paper.</i>
question		<i>Did you drink any coffee today?</i>
conditional		<i>If you see any problem, please let me know.</i>
irrealis	TAM marking, modals	<i>He could have bought any dog.</i>
	inherent irrealis predicates	<i>She is looking for any student who can help her with the organization.</i>
free choice		<i>You can take any cookie.</i>

Without going into any detail with respect to the semantic and pragmatic conditions of each type of contexts illustrated in Table 4.7, we can make use of the concept of veridicality proposed and argued for in F. Zwarts (1995) and Giannakidou (1997, 1998), which accounts for these different uses of adnominal NPIs such as *any* in English. As Giannakidou (2017: 21) puts it: “[a] veridical context is one that allows the speaker to infer the truth of a sentence; a non-veridical context

⁹Work on NPIs goes back to C. L. Baker (1970), Fauconnier (1975), Horn (1972), Ladusaw (1980) who discuss NPIs in English focusing on negation contexts. Amongst others, Giannakidou (1997) and F. Zwarts (1981, 1995) extended the study of NPIs to other polarity domains outside of negation and free choice contexts, also arguing for the distinction between different types of NPIs. For recent overviews on the topic and further references, see e.g. Giannakidou (2017), Hoeksema (2013), Horn & Kato (2000).

is one where truth inference seems to be suspended”. She provides the following example for illustration:

- (174) a. Unfortunately, Mary saw a snake. \Rightarrow Mary saw a snake.
 b. Maybe Mary saw a snake. \nRightarrow Mary saw a snake.
 (Giannakidou 2017: 21)

In example (174a), also in the presence of a “factive” adverbial, the truth value of the utterance (with respect to the discourse universe) can still be assessed which makes the utterance veridical. The utterance in (174b), on the other hand, lies within the scope of the modal *maybe*, which blocks the inference of the truth value of that utterance. This is an example of a non-veridical statement.

Hence, non-veridicality can be taken as the property of an utterance meaning that licenses NPIs. Negation is a special case, as it does not only involve non-veridicality (the truth value of *p* cannot be assessed) but anti-veridicality, which asserts that the proposition does not hold (*p* is not). In addition, Giannakidou (2017: 7) mentions an anti-episodicity constraint: if the utterance makes reference to a single past event, not even negation can license NPIs as, e.g., free choice items:

- (175) a. Any cat hunts mice.
 b. ?Any cat hunted mice.

In (175a), *any* makes the referent nonspecific as a free choice marker, entailing that every referent that belongs to the kind of ‘cat’ can be referred to. Such an interpretation however is only available because the utterance allows for reference to multiple nonspecific events. The utterance in (175b), on the other hand, is tied to a single past event, which blocks the nonspecific free choice interpretation of *any* and makes the sentence infelicitous.

The contexts for nonspecific articles are exactly the same (cf. Section 6.4). They signal that there is no particular referent that is linked to the expression used, but that each potential referent of its kind can be linked to the expression. The distinctive property between adnominal NPIs such as *any* in English and nonspecific articles lies in their distribution. Languages such as English, where the indefinite article *a*, depending on the context, leaves the referent ambiguous between a specific and nonspecific interpretation, do not need to resolve the ambiguity between specific and nonspecific reference. For this reason, we can use the indefinite article *a* or a bare noun instead of the NPI *any* in all contexts that were presented to be relevant to NPIs in Table 4.7:

- (176) a. I did not see **cake** in the fridge. (negation)
 b. I don’t think there was **cake** in the fridge. (neg-raising)
 c. Tabea wrote a dissertation without having **coffee**. (indirect negation)

- d. Tomas failed to write a **paper**. (negative verb)
- e. Did you drink **coffee** today? (question)
- f. If you see a **problem**, please let me know. (conditional)
- g. He could have bought a **dog**. (irrealis)
- h. She is looking for a **student** who can help her with the organization. (irrealis verb)
- i. You can take a cookie. (free choice)

Of course, even though the utterances from Table 4.7 can occur with indefinite articles or bare nouns instead of NPIs with their truth value being the same (Chierchia 2013: 27), there are differences between the interpretations of the sentences in Table 4.7 and in example (176) above. We find differences of two types. Either, using the article or a bare noun, the referent is ambiguous between a specific and nonspecific interpretation in the contexts in which no other explicit linguistic cue for nonspecificity is provided. We see this in (176g) and (176h). Or, NPIs can involve pragmatic effects that are absent in the sentences in (176). Such effects have been linked to exhaustivity (e.g. Chierchia 2013: 27) and alternatives similar to focus (e.g. Chierchia 2013: 34, Giannakidou 1998: 81). What is important for the purposes of the present study is the fact that although NPIs generally seem to be available in the languages of the world in the same contexts as nonspecific articles occur in, the examples in (176) showed that NPIs are not required to be used with nonspecific referents systematically. In other words, with NPIs, a nonspecific referent does not have to be marked as such, whereas it does if the marker in question is a nonspecific article.

As mentioned in the previous section, Ayoreo has a nonspecific article. The following example shows that the nonspecific article *-tique* is not acceptable in specific contexts as in (177a), where the inclusive-specific article *-i* has to be used instead. The nonspecific article is however systematically required in contexts like (177b), where the referent is interpreted as nonspecific.

- (177) a. María pota nona *ata-tique / ata-i uté uje
 María wants accompanies rich-ART:NSPEC.M.SG / rich-ART:INSPEC.M.SG that COMP
 chise dirica.
 meets yesterday
 ‘María wants to marry that rich man whom she met yesterday.’
- b. María pota nona ata-tique / *ata-i mu cama
 María wants accompanies rich-ART:NSPEC.M.SG / rich-ART:INSPEC.M.SG but not.yet
 chimó.
 sees
 ‘María wants to marry a rich man, but she has not yet met him.’
 Ayoreo (Bertinetto 2009: 46-47)

Thus, if there is no systematic need to resolve the ambiguity between specific and nonspecific referents in the noun phrase, markers that can occur to mark the referent as nonspecific are not treated as articles. If referents have to be marked as either specific or nonspecific in a systematic way, I treat such markers as nonspecific articles.

4.5 Summary

In this chapter, I discussed how certain types of articles can be distinguished from similar related markers. The distinction between definite articles and demonstratives was shown to be problematic because of their diachronic connection. I followed earlier proposals from the literature that demonstratives are restricted to pragmatic definite contexts, i.e. to deictic, anaphoric, recognitional, and establishing referents. Definite articles, on the other hand, are used with both pragmatic definite referent types (anaphoric, recognitional, establishing) and semantic definite referent types (situationally unique, contextually unique, bridging). Taken together with anaphoric and recognitional articles, I argued that the minimal criteria for definite articles are to occur in with anaphoric and contextually unique referents. Anaphoric articles must code anaphoric referents and must not occur with deictic or contextually unique referents. Vice versa, weak definite articles occur with contextually unique referents but not with anaphoric ones. Finally, recognitional articles only encode recognitional referents.

The second part of this section examined the similarities and differences between possessive markers and definite articles. In accordance with the conclusions from earlier studies, I argued for a fundamental difference between the distribution of the possessive marker in Udmurt and definite articles in other languages. Even though the possessive marker occurred in all contexts that definite articles need to occur in, all evidence suggested that it does not do so systematically, i.e. the presence of the possessive was not required in the relevant definite contexts. In addition, I briefly mentioned the argument from the literature against a the diachronic development of *POSS* > *ART:DEF* in Semitic, where there seems to be evidence for two separate developments from *DEM* > *ART:DEF* and *DEM* > *POSS* instead. I presented Indonesian as another case where a possessive is used as definite marker, showing that the possessive is not a possessive in the strict sense, as it serves a number of other syntactic, semantic, and pragmatic functions. Against this background, I concluded that it may be necessary to revisit the crosslinguistic evidence suggesting possessive markers as a direct source of definite articles.

In the third part of this section, I discussed how articles in the indefinite domain can be distinguished from the numeral ‘one’. I argued that numerals which can occur in presentational uses should not be considered articles, because this functional extension of the numeral ‘one’ seems to be generally available across languages. I argued that a marker occurring only with discourse-

prominent specific and nonspecific referents is a special type of indefinite article, which I call presentational article. If a marker is systematically used to encode specificity independently of the discourse prominence of the referent, it is regarded as an exclusive-specific article. If it marks only nonspecific referents, it corresponds to a nonspecific article.

The last part discussed the distinction between nonspecific articles and negative polarity items (NPIs), since adnominal NPIs have similar functions to the markers that I treat as nonspecific articles. I argued that the difference between nonspecific articles and NPIs does not lie in the contexts that they occur in but in the systematicity of their distribution with nonspecific referents. NPIs are not systematically required to mark referents as nonspecific, and if they occur, they may have additional pragmatic effects that nonspecific articles lack. Nonspecific articles occur systematically to encode a referent as nonspecific; in contrast to NPIs, they are required in such contexts.

Chapter 5

Articles in the definite domain

This chapter gives an overview of definite (Section 5.1), anaphoric (Section 5.2), weak definite (Section 5.3), and recognitional (Section 5.4) articles. For each type, I provide several examples from the world's languages. In addition to the detailed discussion of the different article types, this chapter makes two important observations concerning the properties of definite articles. I show in Section 5.1.3 that the coding of deictic referents is not a necessary criterion for definite articles. Rather, their occurrence in deictic contexts can be accounted by the fact that most definite articles originate from demonstratives that are used in such contexts as well. As I show for Kaqchikel, definite articles with sources other than spatial demonstratives need not extend their functional domain to the coding of deictic referents. In addition, I discuss the occurrence of definite articles together with absolutely unique referents (cf. Section 5.1.4) and with proper nouns (cf. Section 5.1.5), showing that definite articles can but do not have to occur in such contexts, without any noticeable crosslinguistic trend.

5.1 Definite articles

This section deals with definite articles which are articles that occur systematically with (at least) anaphoric and contextually unique referents, as defined in Section 3.3.1. In addition to those two contexts, I expect definite articles to systematically express bridging, establishing, and recognitional referents as well. However, these types of contexts are not always well-documented, which is why I treat an article as definite on the basis of its systematic presence with anaphoric and contextually unique referents. This section first zooms in on the definite articles in Kaqchikel (Mayan, Guatemala) in Section 5.1.1 and Mokpe (Bantu, Cameroon) in Section 5.1.2. The use of definite articles with deictic referents is discussed in Section 5.1.1. The compatibility of definite articles with absolutely unique referents and proper nouns is then addressed in Sections 5.1.4 and 5.1.5, respectively.

5.1.1 Kaqchikel

Kaqchikel is a Mayan language spoken in Guatemala by approximately 500 000 speakers.¹ The language has an indefinite article, *jun*, and a marker *ri* that is usually treated as a definite article (R. M. Brown et al. 2006, Chonay Chonay 2006). This section is concerned with the definite article; I will show that *ri* is systematically used to mark anaphoric, recognitional, establishing, bridging, and situationally as well as contextually unique referents. Moreover, it does not occur with specific or nonspecific referents, for which Kaqchikel uses the indefinite article *jun*.

Example (178) shows an anaphoric context. In (178b), the article *ri* is required to mark an anaphoric referent in a short conversation between A and B. The expression *ixoq* ‘woman’, uttered by speaker B, refers back to the referent of *jun ixoq* uttered by speaker A in (178a).

- (178) a. tak’ xinapon chuwe choch xintz’at [jun ixoq] ruk’in kayi aq’ala
when arrive.PST.1SG to.1SG house see.PST.1SG ART.INDEF woman with two child
ech’okol chuchi qaq’
seated close fire
‘When I came home, I saw a woman with two children sitting in my kitchen.’
- b. awetaman ruwäch [*(ri) ixoq]?
know.PST.2SG eyes.3SG ART:DEF woman
‘Did you know the woman?’
Kaqchikel (primary data)

Another anaphoric context in which the definite article occurs with the noun is illustrated in (179). We see that the definite article *ri* is also required with the abstract noun *ajob’ab’al* ‘love’ to mark anaphoric reference.

- (179) ri amistad xok ajob’ab’al. [*(ri) ajob’ab’al] xok etzelanik.
ART:DEF friendship turn.PST.3SG love ART:DEF love turn.PST.3SG hatred
‘The friendship turned into love. The love turned into hatred.’
Kaqchikel (primary data)

Definite articles are also required to encode rel-bridging referents. Like anaphoric referents, rel-bridging referents have an antecedent that an anaphoric relation can be established with. Example (180) illustrates the use of the definite article in Kaqchikel with rel-bridging referents. Here, the reference to preceding *ch’ich* ‘car/taxi’ enables the definite marking of *chofer* ‘driver’ because ‘driver’ is a relational concept, allowing for an anaphoric relation between ‘driver’ and ‘car’. This results in the unambiguous identifiability of the bridging referent, which is obligatorily marked by the definite article.²

¹This figure is taken from the Endangered Languages Project (<http://www.endangeredlanguages.com/lang/8089>).

²I use the notions “definite” and “identifiable” interchangeably.

- (180) k'ochi xinchob jun ch'ich' chila pa centro. [***(ri)** chofer]
 have.to.PST.1SG take ART:INDEF car there in centre ART:DEF driver
 xup'ij chwa chiri ramajri manaqchik ch'ich'.
 tell.PST.S:3SG.O:1SG to.1SG COMP this.hour is.not car
 'I had to take a taxi to the centre. The driver told me that there were no cars at that hour.'
 Kaqchikel (primary data)

Still within the domain of pragmatic definite uses, definite articles are expected to mark establishing referents, i.e. mark a referent as identifiable so that the hearer can establish identifiability using the information provided by the speaker in the immediately following discourse segment. Often, a construction with a relative clause that follows the noun is used, as in examples (181) and (182). Both referents of *achik* 'dream' and *achin* 'man' are marked as identifiable by the use of the definite article *ri*, and are made identifiable in the following relative clause, establishing their identity.

- (181) tatzijoj chuwe [***(ri)** achik'] xab'en rijunkän aq'a.
 tell.IMP to.1SG ART:DEF dream make.PST.2SG last night
 'Tell me about the dream you had last night.'
 Kaqchikel (primary data)

- (182) [***(ri)** achin] xuk'ul Mario iwir jari xoyompe k'aba.
 ART:DEF man meet.PST.3SG Mario yesterday be call.PST.3SG recently
 'The man that Mario met yesterday called not long ago.'
 Kaqchikel (primary data)

Similarly, the definite article in Kaqchikel can also be used to mark a recognitional referent, signaling to the hearer that the referent is identifiable based on shared knowledge or experience, but not from the current discourse situation. An example of a recognitional referent is given in (183), where the referent of *xintäj* 'fruit' is marked by *ri* as identifiable because of shared experience prior to the current discourse situation. The context of (183) is the following: Speaker A buys a fruit for speaker B. On the next day, speaker B eats the fruit and tells speaker B to thank her, starting the conversation with the utterance in (183).

- (183) xintäj [***(ri)** jiq'ob'al].
 eat.PST.1SG ART:DEF fruit
 'I ate the fruit (that you bought yesterday).'

That the definite article is used with situationally unique referents as well is shown in example (184). Here, the referent of *je'bel* 'food' is identifiable only based on the immediate discourse situation; it is not deictic, though, because at the time that (184) is uttered, the physical object

linked to the referent is no longer present in the situation. We see that use of the definite article is required.

- (184) [***(ri)** qutun] jeb'el xub'en.
ART:DEF food delicious result.PST
 [Context: The speaker complements the hearer on the food that the hearer had prepared and that both just finished eating.]
 'The food was delicious.'
 Kaqchikel (primary data)

The definite article is further required to mark contextually unique referents. A contextually unique referent is identifiable by both the speaker and the hearer because a larger context can be evoked in which the referent is constructed as unique or as the only relevant of its kind. Example (185) allows for the unambiguous identification of the referent of *k'amor-b'äl* 'president', since the concept of 'president' is unambiguously linked to the concept of 'country', which is evoked by the larger context of the discourse situation being placed in Guatemala. As can be seen in (185), the definite article must be used.³

- (185) [***(ri)** k'amor-b'äl] jani rusamaj wakami.
ART:DEF guide-way much work.3SG now
 'At the moment, the President (of Guatemala) has a lot of work.'
 Kaqchikel (primary data)

Another example of a contextually unique referent is presented in (186). Accordingly, both the speaker and the hearer know based on their world knowledge that there is maximally one hospital in the town of the discourse situation (in this case, Patzún in Guatemala). Therefore, the referent of *aq'omanel-jay* 'hospital' is contextually unique and marked as such by the definite article.

- (186) ankuchi k'obi [***(ri)** aq'omanel-jay] (chin re tinamet re)?
 where **EXIST ART:DEF** doctor-house in **DEM** town **DEM**
 'Where is the hospital in this town?'
 Kaqchikel (primary data)

U-bridging contexts form a subtype of contextually unique referents in which a definite article is expected as well. In (187), we see a part-whole relation between *jay* 'house' and *ruwi* 'roof'. The referent of *jay* 'house' makes the following referent of *ruwi* 'roof' identifiable and thus definite, because the frame of the concept 'house' enables us to construct the referent of 'roof' as contextually unique. The same holds for the relation between *k'ulbik* 'wedding' and *xten nik'ule* 'bride' in (188). In both cases, the definite article has to be used with the u-bridging referents.

³'President' is a typical example of a functional concept in the approach of Löbner (1985, 2011), and it is this one-to-one relation between 'country' and 'president' that makes the referent of 'president' contextually unique.

- (187) xk'atz'enkän jun jay tziaqchik chila pa tinamet. [***(ri)** ruwi] jmul
 see.PAST.1PL ART:INDEF house old there in village ART:DEF roof completely
 manutzachik.
 non.functioning
 'We saw an old house there in the village. The roof was completely damaged.'
 Kaqchikel (primary data)
- (188) xibe pa jun k'ulbik. [***(ri)** xten nik'ule] jab'el (ok) ri
 go.PST.1SG to ART:INDEF wedding ART:DEF lady get.married pretty EMPH POSS:3SG
 rutziaq.
 dress
 'I went to a wedding. The bride had a pretty dress.'
 Kaqchikel (primary data)

The previous examples thus showed the systematic occurrence of the definite article in Kaqchikel with all relevant types of definite referents. I will now show that *ri* is indeed restricted to definite contexts and does not occur with non-identifiable referents, which are marked by the indefinite article *jun* in Kaqchikel.

In examples (189), (190), and (191), *ri* cannot be used in the intended meaning. These examples show specific referents, which correspond to particular referents that are linked to the expression used without being unambiguously identifiable by all discourse participants. This is what we see for the referents of *jay* 'house', *lugar* 'place', and *chkop* 'animal', which the hearer cannot identify. To mark that the hearer (and in the case of (191), also the hearer) lacks sufficient knowledge to unambiguously identify those referents, the speaker uses the indefinite article *jun*.

- (189) xk'atz'enkän [**jun** jay tziaqchik] chila pa tinamet.
 see.PST.1PL ART:INDEF house old there in village
 'We saw an old house in there in the village.'
 Kaqchikel (primary data)
- (190) Marta xutzijoj chuwe chi k'o [**jun** lugar janila interesante].
 Marta tell.PST.S:3SG.O:1SG to.me COMP EXIST ART:INDEF place very interesting
 'Marta told me about a very interesting place.'
 Kaqchikel (primary data)
- (191) tatzu xachob' [**jun** chkop].
 look.IMP catch.PST.2SG ART:INDEF animal
 'Look! You caught an animal!'
 Kaqchikel (primary data)

Examples (192), (193), and (194) below show that the same holds for nonspecific referents. Nonspecific referents correspond to a single but no particular referent of its kind, and they are marked

by the indefinite article *jun* as well. Note that, again, the definite article *ri* cannot occur with *azib'ab'al* ‘pen’, *ch'ich'* ‘car’, and *b'anoy-qutun* ‘cook’ with the intended meaning.

- (192) *k'o [jun atzib'ab'al]*.
 EXIST ART:INDEF pen
 ‘Do you have a pen? (Any pen will do.)’ literally: ‘Is there a pen?’
 Kaqchikel (primary data)
- (193) *ninwajo [jun ch'ich']*.
 want.1SG ART:INDEF car
 ‘I want to have a (any) car.’
 Kaqchikel (primary data)
- (194) *nk'atzin chuwe [jun b'anoy-qutun]*.
 need.1SG to.me ART:INDEF prepare-food
 ‘I need a (any) cook.’
 Kaqchikel (primary data)

Thus, we saw that *ri* is a definite article, occurring with all relevant definite referential functions while being absent in specific and nonspecific contexts. The contexts of occurrence of *ri* are summarized in Table 5.1 below; the numbers in brackets refer to the respective examples. As will be shown in Section 5.1.3, the definite article in Kaqchikel cannot be used with deictic referents; examples of its compatibility with absolutely unique referents are given in Section 5.1.4.

Table 5.1: The distribution of *ri* in Kaqchikel

referential functions	use of <i>ri</i>
R_{deictic}	✗ (216), (217)
R_{anaph}	✓ (178), (179)
R_{recog}	✓ (183)
R_{establ}	✓ (181), (182)
$R_{\text{rel-bridge}}$	✓ (180)
$R_{\text{sit.unique}}$	✓ (184)
$R_{\text{cont.unique}}$	✓ (185), (186)
$R_{\text{u-bridge}}$	✓ (187), (188)
$R_{\text{abs.unique}}$	✓ (216),(217)
R_{specific}	✗ (189), (190), (191)
$R_{\text{nonspecific}}$	✗ (192), (193), (194)

5.1.2 Mokpe

Another example of a definite article can be found in Mokpe, a Narrow Bantu language spoken in the area of Buea, Cameroon. Mokpe has approximately 35 000 speakers (Atindogbe 2013: 5). The definite article in Mokpe is interesting because of its exponent, which can be segmental or tonal depending on the class of the noun it occurs with. Mokpe does not have any other articles; if the referent of the noun is specific or nonspecific, it does not require marking as such and the bare noun is used.

In his grammatical sketch of Mokpe, Atindogbe (2013: 11) mentions articles only briefly, noting that “the indefinite and definite articles are rendered by the low and the high tone respectively”. In this section, I show that the definite article can indeed have the form of a high tone on the noun. However, I will show that the definite article also has a different segmental exponent, which is not mentioned explicitly in Atindogbe (2013), even though the marker does occur in certain examples. Another difference between Atindogbe’s and my analysis is that I do not treat a noun without the definite article (either as a high tone or as a segmental marker) as a noun with an indefinite article. Such a form rather corresponds to the bare noun in all non-definite contexts.

As is commonly found in Bantu, Mokpe has an elaborate noun class system, where the gender and number values occur as fused markers on the noun itself, on other elements in the noun phrase, and as agreement markers on the verb. Table 5.2 shows the article exponents for each noun class. The class distinction is adapted from Atindogbe (2013); all examples in this section and the analysis of the article system are based on primary data collected in Buea, Cameroon. Classes 1–10 correspond to singular-plural pairs. For each class, two nouns are shown with a high tone (´) and a low tone (`) on their first tone-bearing unit in their singular and plural forms.⁴ Nouns of classes 14 and 19 contain singular nouns; since these nouns have (most of) their plural forms in class 10, no corresponding plural classes are indicated for class 14 and 19.

For nouns of class 1 and 9, the article is realized as the segmental marker *è*.⁵ For nouns of the other classes, the article does not surface as a segmental marker but as a tonal process on the noun, which can be described as a high tone on the first tone-bearing unit of the noun. With nouns of the classes that require the tonal exponent of the definite article, it surfaces as a high tone on the first tone-bearing unit of the noun if the latter has a lexical low tone in this position (first row of examples). If the first tone-bearing unit of the noun is already lexically specified as having a high tone, the tonal article does not affect the tones of the noun (second row of examples). Only for the nouns in classes 5 and 6, a phonological effect of the tonal article could be observed with nouns

⁴My consultants and I are not aware of examples with initial high tones for nouns of class 19.

⁵Note that nouns of class 10 occur with the prenominal marker *í* independently from whether the referent expressed is definite or not. Thus, *í*, unlike *è* in class 1 and 9, does encode referentiality.

Table 5.2: Paradigm of the definite article in Mokpe

class (SG)	noun	article+noun	class (PL)	noun	article+noun	
1	mòtò ɲmána	è mòtò è ɲmána	2	wàtò wána	wátò wána	'person' 'child'
3	mólélí mólèngò	mólélí mólèngò	4	mèlélí mélèngò	mélélí mélèngò	'food' 'sheep'
5	likàlà líkpà	líkàlà lí:kpà	6	màkàlà mákpà	mákàlà má:kpà	'bridge' 'bag'
7	ètángùlè élélé	ètángùlè élélé	8	βètángùlè βélélé	βètángùlè βélélé	'lizard' 'morning'
9	mbèzà mbóli	è mbèzà è mbóli	10	í mbèzà í mbóli	í mbèzà í mbóli	'boy' 'goat'
14	wòzǒ wóɲgó	wózǒ wóɲgó			/	'face' 'fear'
19	ìnòní dzùngù	ìnòní dzùngù			/	'bird' 'mosquito'

that already have an initial high tone. This effect is represented as vowel-lengthening here.⁶ It is very likely that the tonal article exponent represents an older article system. It is more pervasive and the only marking for definite referents mentioned in *Atindogbe (2013)*.

The fact that a tonal and a segmental marker can mark nouns of different classes as definite is a strong argument for the form-independence of articles. As we will see in this section, the tonal process has the same referential function as the segmental marker *è*. Therefore, I treat them as equal exponents of the definite article in Mokpe.

Although there is no reason that would speak against referential markers being expressed by tones, tonal article exponents are crosslinguistically rare. The only other reported case that I am aware of is the definite article in Ewondo (*Lyons 1999: 65-66*), a Bantu language closely related to Mokpe. Tonal articles probably originate from short(ened) segmental prenominal morphemes that are phonetically reduced to the nonsegmental material, which then associates to the first tone-bearing unit of the noun. Other similar processes of phonetic reduction in Mokpe make such a scenario plausible.⁷ For instance, phonetic reduction is pervasive with prepositions and

⁶In order to fully account for the tonal effects of the article in Mokpe, a detailed phonetic and phonological analysis would be necessary. I present the system in a simplified way here, without discussing contour tones that are affected in a slightly different way by the tonal article.

⁷The phonological assimilation of preposed articles is common across the world's language, a well-known example being the definite article *le/la* in French occurring as shortened *l'* before vowel-initial nouns.

other function words that often only consist of a single vowel and can be reduced to a tone which associates to the first tone-bearing unit of the following word.

That the definite article is used in anaphoric contexts is shown in (195) and (196). In both examples, the referents of *móláná* ‘woman’ and *málúwá* ‘water’ are identifiable because they were mentioned in the preceding utterance, which requires the use of the definite article. In (195b), we see the article exponent *è* with a noun of class 1.

- (195) a. *è* *βóndá* na *dzèlí*, n-*éní* **móláná** à *líj-ó* *kízèni* nà *βánà* *βá*
 ART:DEF time I came 1SG-see woman.CL1 CL1 sit-in kitchen with children.CL2 CL2
wàké.
 two
 ‘When I came (home), I saw a woman with two children sitting in the kitchen.’
- b. *ò* *mà wíjá* [**è*] *móláná*?
 you PST know ART:DEF woman.CL1
 ‘Did you know that woman?’
 Mokpe (primary data)

The noun *málúwá* in (196) takes the tonal exponent of the definite article. When the noun *málúwá* ‘water’ is first mentioned in (196a), evoking the concept of water, its first tone-bearing unit has a low tone. Used anaphorically in (196c), the definite article surfaces as a high tone on the noun, realized as *málúwá*.

- (196) a. nà *βéja* *è* *nízá* **málúwá**.
 I feel ART:DEF thirst water.CL6
 ‘I am thirsty for (water).’
- b. **málúwá** *má βél-ǒ* *kízèni*.
 water.CL6.ART:DEF CL6 COP-in kitchen
 ‘(The) water is in the kitchen.’
- c. *džǎ té*, nà *zèni* **málúwá** / **máluwa*.
 come please I see.NEG water.CL6.ART:DEF / water.CL6
 ‘Please come, I cannot find the water.’
 Mokpe (primary data)

Examples (197) and (198) below show the use of the definite article in Mokpe with establishing referents. In those two examples, the referents marked as definite are made identifiable to the hearer in the immediately following discourse segments. The establishing information is expressed by a relative clause in (197) and by a syntactically independent clause in (198).

- (197) [**è*] *mòtò*] *Ngúdú* á *mùnmá* *wúnùwù* à *wélê*.
 ART:DEF person.CL1 Ngudu 3SG meet.PST last.night 3SG call
 ‘The man Ngudu met last night (just) called.’
 Mokpe (primary data)

- (198) **ɲméndì** / ***ɲmèndì** mó wéli ndí dzòkě? wúwèlèlè ó mbèngà wàkálà li
news.CL3.ART:DEF / news.CL3 CL3 COP true leave.3PL LOC down white.CL2 INF
 dzǎ li dzě̀nè ánú?
 come INF see.1PL here
 ‘Are the rumors true? Are they coming to visit us from Europe?’
 Mokpe (primary data)

A definite article has to be used with nouns in rel-bridging contexts as well. Example (199) illustrates this use of the definite article in Mokpe. In (199), the referent of *mòtìlèlì* ‘writer’ is identifiable because it is a functional concept with a relation to the concept of ‘book’. Thus, the referent of *mòtìlèlì* ‘writer’ can establish an anaphoric relation to the previously mentioned referent of *kàtì* ‘book’. Since the noun *mòtìlèlì* ‘writer’ belongs to class 1, the article surfaces as è.

- (199) mòkákè àndí kàtì. [***(è)** mòtìlèlì] à βèlì ndí mòtá βéndì.
 Mokake buy.3SG book **ART:DEF** writer.CL1 CL1 COP person.AM French
 ‘Mokake bought a book; the author is French.’
 Mokpe (primary data)

The following examples show that the definite article in Mokpe also occurs with semantic definite referents. Example (200), the counterpart of (185) from Kaqchikel, illustrates the use of the definite article in such contexts. Uttered in Cameroon, the referent of ‘president’ can be constructed as contextually unique with the country as the larger context of the discourse situation. This is why the referent of *mòkànèlì* ‘leader (president)’ is unambiguously identifiable and requires to be marked by the article. Moreover, the utterance in (200) contains the noun *ndzùmá* ‘fight, strike’ which is equally marked as definite. The referent of *ndzùmá* ‘fight, strike’ is a recognitional referent; both the speaker and the hearer know about the strikes against the repressions from the francophone government towards the anglophone population in Cameroon in 2017, and the strikes have been part of different previous conversations and experiences, without having been mentioned before in the current discourse situation.

- (200) [***(è)** mòkànèlì] à zóβ-è βóndá jà [***(è)** ndzùmá].
ART:DEF leader.CL1 CL1 NEG-have time for **ART:DEF** fight.CL9
 ‘The president does not care about the strike.’
 literally: ‘The leader does not have time for the fight.’
 Mokpe (primary data)

Example (201) shows the Mokpe counterpart of example (187) from Kaqchikel with a part-whole relation between a house and its roof, i.e. a typical u-bridging context. Again, we see that the definite article is used together with the expression of *likàndò* ‘roof’. Since this noun belongs to class 5, it requires the high tone exponent.

- (201) *ì mènέ ημάνέ βòβá ndáwò ó mbówà. líkândò / *líkândò lí mà*
 we see.PST.1PL small ugly house in village roof.CL5.ART:DEF / roof.CL5 CL5 PST
kúwú lí βáηgèjà.
 finish INF get.damaged
 ‘We saw a small ugly house in the village. The roof had been damaged.’
 Mokpe (primary data)

Example (202) shows another bridging example.⁸ By mentioning the referent of *è mbówà jèní* ‘his village’ in (202a), the concept of ‘village’ makes the people living in that village, and the “typical” landscape surrounding the village (the river and mountains) the only salient referents of their kinds, i.e. they are contextually unique. Therefore, the referents of *wátò* ‘people’, *mòlélí* ‘food’, *mòzô* ‘river’, and *βákó* ‘mountain’ are definite and marked as such by the definite article in (202b). The nouns *wátò*, *mòlélí*, and *mòzô* occur with a high tone on their first tone-bearing unit, while the noun *βákó* requires the segmental article exponent *è*. Example (202c) shows that the absence of the definite article in this context is not felicitous.

- (202) a. *mòkákè à lijà ó lùwálá nánḡê, ndí ó ηmémá ηméní à lijúní è*
 Mokake 3SG live in Douala now but in heart.CL3 POSS.CL3 3SG love ART:DEF
mbówà jèní ...
 village.CL9 POSS.CL9 ...
 ‘Mokake lives in Douala now, but in his heart, he still loves his village ...’
- b. ... *wátò, mòlélí, mòzô nà [*è]*
 ... people.CL2.ART:DEF food.CL3.ART:DEF river.CL3.ART:DEF and ART:DEF
βákó].
 mountain.CL9
 ‘... The people, the food, the river, and the mountain.’
- c. *... *wátò, mòlélí, mòzô nà βákó.*
 ... people.CL2 food.CL3 river.CL3 and mountain.CL9
 intended: ‘... The people, the food, the river, and the mountain.’
 Mokpe (primary data)

In order to qualify as a definite article, the article in Mokpe also needs to be infelicitous with specific and nonspecific referents. This is what we see in the following examples. Examples (203) and (204) show specific referents which are expressed by the bare nouns *ndáwò* ‘house’ and *pámà* ‘animal’. The definite article cannot be used with the intended meaning, i.e. marking the referents as non-identifiable.

- (203) *ì mènέ ημάνέ βòβá [(*è ndáwò) ó mbówà.*
 we see.PST.1PL small ugly [ART:DEF house.CL9] in village
 ‘We saw a small ugly house in the village.’
 Mokpe (primary data)

⁸This is probably a case of bridging which is ambiguous between unique and relational bridging.

(204) óηgó! è ñùndèlì è kǒwí [(*è) ñàmà].
 look.IMP ART:DEF trap.CL9 CL9 catch ART:DEF animal.CL9
 ‘Look! The trap caught an animal.’
 Mokpe (primary data)

Two examples of nonspecific referents are given in (205) and (206). Here, no particular referent is evoked by the referential expression. We see that the definite article cannot occur in such contexts with the intended meanings either.

(205) nà àzrà [(*è) mòtò] wà mòkáw à wí lííẏròηgò nómà ná mà wíjé.
 I need ART:DEF person.CL1 AM hunt 3SG know hunting like I PST know
 ‘I need a (any) hunter who can hunt like I used to do.’
 Mokpe (primary data)

(206) wúpá wókó nà óβà mòtówà / *mótówà.
 day one I have car.CL3 / car.CL3.ART:DEF
 ‘One day I will have a (any) car.’
 Mokpe (primary data)

Thus, this section established the two markers *è* and the high tone (H) on the first tone-bearing unit of the noun as exponents of the definite article in Mokpe. Table 5.3 summarizes the contexts of occurrence of the definite article; the numbers in brackets refer to the respective examples. Its occurrence with deictic referents is marked in brackets because the definite article in Mokpe is compatible with deictic referents (cf. Section 5.1.3). Its occurrence with absolutely unique referents is shown in Section 5.1.4.

Table 5.3: The distribution of *è*/H in Mokpe

referential functions	use of <i>è</i> /H
R _{deictic}	(✓) (207), (208), (208)
R _{anaph}	✓ (195), (196)
R _{recog}	✓ (200)
R _{establ}	✓ (197), (198)
R _{rel-bridge}	✓ (199)
R _{sit.unique}	✓ (207), (208), (208)
R _{cont.unique}	✓ (200)
R _{u-bridge}	✓ (201), (202)
R _{abs.unique}	✓ (221)
R _{specific}	✗ (203), (204)
R _{nonspecific}	✗ (205), (206)

5.1.3 Definite articles with deictic referents

Although definite articles are often compatible with deictic referents because of the high degree of overlap between deictic and situationally unique referents, they do not have to occur in uses. This section discusses the occurrence of definite articles with deictic referents in more detail, showing that definite articles across languages vary regarding the behaviour in deictic contexts.

We can distinguish three main patterns: definite articles can either occur with deictic referents that are not marked as such otherwise, which would suggest that they do in fact encode the deictic referent in such cases. Or, they can occur with deictic referents in addition to a demonstrative, in which case it seems more plausible to assume that the definite article does not mark the deictic referent as such but is compatible with it. The third pattern that we find does not allow the use of a definite article with deictic referents, in that case, the definite article is clearly incompatible with deictic referents.

Definite articles like *the* in English belong to the first type, being compatible with the expression of deictic referents. One could argue that the reason for this compatibility is their diachronic development from a demonstrative with spatial deictic functions. In such a scenario, the definite article may simply retain the function of marking deictic referents as definite. Because the contexts of a deictic and a situationally unique referent overlap to such a great extent, it may not be possible to disambiguate the two types of referents in deictic contexts, as was already noted in Hawkins (1978) and Himmelmann (1997).

To show that this property of definite articles is not restricted to the well-known Indo-European examples, consider the definite article in Mokpe presented in the previous section, which is compatible with deictic referents as well. Examples (207) and (208) show that the definite article can be used as the only linguistic marker of the deictic referent and its unambiguous identifiability. In both examples, the speaker points to the objects in the discourse situation which are associated with the referents, which requires the referent to be interpreted, at least partially, as a deictic referent.

(207) [è téwèlì] è gbèjána wí wèwámbo.

ART:DEF table.CL9 CL9 made of wood

[Context: The speaker S points at the only table in the room, visible to both S and the hearer H.]

‘The table is made of wood.’

Mokpe (primary data)

- (208) óṅgó! [è ṅùndèli] è kǒwí òàmà.
 look.IMP ART:DEF trap.CL9 CL9 catch animal
 [Context: The speaker S points at the trap visible to both S and the hearer H.]
 ‘Look! The trap caught an animal.’
 Mokpe (primary data)

In example (209), the definite article is shown to be functionally equivalent to the demonstrative marker. It is still possible, though, that the definite article marks the referent as situationally unique, while the demonstrative as deictic, only that the context cannot disambiguate between these two types of referential functions.

- (209) [òno / è ṅmánà] á má kòkà liwòtèjà è ßóndá ná mà ṅmènê.
 DEM:PROX / ART:DEF child 3SG RES grown since ART:DEF time I PST see.O:3SG
 [Context: Speaker S and hearer H are in a room with one child, visible to both of them.]
 ‘This child has grown since the last time I saw her.’
 Mokpe (primary data)

A more fine-grained distinction of such contexts and potential differences regarding the use of demonstratives and definite articles would go beyond the scope of the present study. The important point that the examples from Mokpe showed is that the definite article can at least be compatible with the interpretation of a referent as a deictic one in contexts in which the definite article is the only referential marker of the nominal expression.

This kind of compatibility between the definite article and a referent that is deictic and situationally unique is attested in other languages from the sample as well, even though clear examples of such contexts seem very rare. In all examples, the referent also seems to be the only referent of its kind in the discourse situation, meaning that the referent is always both deictic and situationally unique. Taken together, this suggests that even though the definite article may be compatible with deictic uses, it does not mark referents as deictic but as situationally unique (as was argued in Section 4.1). Two other examples for the use of the definite article with situationally unique and deictic referents in the absence of other markers are shown in (210) and (211) for Buwal (Afro-Asiatic, Cameroon) and Nuuhchahnulth (Wakashan, USA), respectively.

- (210) Ma-ṅga [teked anta] vayay?
 REL-break calabash ART:DEF who
 [The broken calabash is present in the room.]
 ‘Who broke the calabash?’
 Buwal (Viljoen 2013: 239)

- (211) maakuk-mit-ḥ č'upč'upšuml-ʔii?
 buy-PST-3.Q sweater-ART:DEF
 'Did he buy the sweater?'

[Context: You know that Ken is planning to buy a sweater for his mother. You see a new sweater on the couch, and want to know if Ken bought it.]

Nuuchahnulth (Wojdak 2008: 80)

Definite articles can also be compatible with deictic referents if they co-occur with demonstratives. In such cases, the demonstrative marks the referent as deictic and the definite article as situationally unique. Examples (212), (213), and (214) show this type of co-occurrence pattern in Bullom So (Atlantic Congo, Guinea & Sierra Leone), Dime (South Omotic, Ethiopia), and Ulwa (Misumalpan, Nicaragua). This co-occurrence pattern between definite articles and demonstratives is rather common across languages; there are many more examples from the sample. However, it is not always clear from the sources whether the demonstrative really marks the referent as deictic.

- (212) [wìs ìcé wǝ̀nǝ̀] lèk b̀ùl k̀ò b̀ì.
 animal ART:DEF DEM horn one PRO have
 'This animal had one horn.'

Bullom So (Childs 2011: 78)

- (213) [sinú ʔiyy-ís] yíd-i-n mǝng-ís-m.
 this.M man-ART:DEF touch-PF-3 gourd-ART:DEF-ACC
 '[Who touched this gourd?] This man touched the gourd.'

Dime (Seyoum 2008: 43)

- (214) [Yaka pan-ka ya] yam-ka daya-p-i sak-pa-h.
 that stick-3SG ART:DEF good-ADJ lean-CL-SS stand-SS stand-CL-2SG.IMPER
 'Stand that stick up at a good angle.'

Ulwa (Koontz-Garboden 2009: 458)

Thus, definite articles are clearly at least compatible with the interpretation of a definite referent as deictic. However, this does not necessarily have to be the case. For instance, the definite article *ri* in Kaqchikel introduced in Section 5.1.1 cannot occur with deictic referents. Consider the two examples in (215), both of which contain an utterance with the referent *ulew* 'ground' marked by a demonstrative in (215a) and by the definite article in (215b). The interpretation of the referent of *ulew* in these utterances necessarily differs. With the demonstrative in (215a), it is interpreted as deictic, as is indicated by the translation. On the other hand, the noun *ulew* 'ground' with the definite article in (215b) only allows for a non-deictic referent; it cannot be accompanied by a pointing gesture and, without further context, corresponds to a generic statement.

- (215) a. [la ulew] tz'il.
 DEM ground dirty
 'The ground (here) is dirty.'
- b. [ri ulew] tz'il.
 ART.DEF ground dirty
 'The ground (in general) is dirty.'
- Kaqchikel (primary data)

Examples (216) and (217) below illustrate a similar effect for the use of the definite article with the nouns *q'ij* 'sun' and *ik'* 'moon'. The versions with the demonstrative in (216a) and (217a) are interpreted deictically, which makes them only felicitous if the objects 'sun' and 'moon' are visible to the hearer and the speaker in the moment of utterance. In case the nouns are marked by the definite article as in (216b) and (217b), the referents of *q'ij* 'sun' and *ik'* 'moon' can no longer be deictic. The only felicitous interpretation of the definite referents is as absolutely unique. As a consequence of this incompatibility of the definite article and a deictic referent, the utterances in (216b) and (217b) are only felicitous if the objects 'sun' and 'moon' are not directly visible neither to the speaker nor to the hearer.

- (216) a. man tatzu [la q'ij].
 NEG look.IMP DEM sun
 'Don't look at the sun.'
 (the sun being directly visible, e.g. from outside of a house)
- b. man tatzu [ri q'ij].
 NEG look.IMP ART:DEF sun
 'Don't look at the sun.'
 (the sun not being visible, e.g. from inside of a house)
 Kaqchikel (primary data)
- (217) a. [la ik'] jani nqalaj chpam re aq'a re.
 DEM moon very be.visible.3SG in DEM night DEM
 'The moon is very bright tonight.'
 (the moon being directly visible, e.g. from outside of a house)
- b. [ri ik'] jani nqalaj chpam re aq'a re.
 DEF.ART moon very be.visible.3SG in DEM night DEM
 'The moon is very bright tonight.'
 (the moon not being visible, e.g. from inside of a house)
 Kaqchikel (primary data)

The incompatibility of the definite article *ri* and deictic referents is likely due to the diachronic source of the article, which is the anaphoric marker *ri ... ri*. The Kaqchikel system of demonstra-

tives including the anaphoric marker is given in Table 5.4 based on R. M. Brown et al. (2006: 25) and Chonay Chonay (2006: 104-105).

Table 5.4: Kaqchikel demonstratives

proximal	<i>re</i> noun <i>re</i>
distal	<i>la</i> noun (<i>la</i>)
anaphoric	<i>ri</i> noun (<i>ri</i>)

For distal *la ... la* and anaphoric *ri ... ri*, the second postposed part is often omitted. The form *la* of the distal demonstrative used in (215a), (216a), and (217a) corresponds to this reduced form. We can thus assume that the anaphoric marker went through this process of shortening in becoming a definite article, since its simple preposed form *ri* corresponds to what I treat as the definite article *ri*. Although *ri* was shown to occur with anaphoric referents in Section 5.1.1, certain contexts still require the full anaphoric form *ri ... ri* and do not allow for the short form *ri*. This happens when the antecedent almost immediately precedes the anaphoric referent, and when both are uttered by the same speaker. Two examples are given in (218) and (219). Note that also in English, the demonstrative *that* is more appropriate than the definite article in these contexts.

(218) a Lu xiroyoj jun b'eychik mambajota yich'o ruk'in [ri vinaq
 PROP Pedro call.PST.S:3SG.O:1SG one way want.NEG.1SG talk with.3SG ANA person
 *(ri)].

ANA

'Pedro called again. I don't want to talk to that person.'

Kaqchikel (primary data)

(219) k'ab'a a Lu xuroyoj ri ru vecino loman chuk'a
 recently PROP Pedro call.PST.S:3SG.O:3SG ART:DEF POSS:3SG neighbour while also
 ink'o yin chila' [ri achin *(ri)] jabey xqa chinwach.

be.1SG I there ANA man ANA nice came in.front

'Recently Pedro invited his neighbour while I was at his house. I liked that man.'

Kaqchikel (primary data)

We can hence assume that the definite article *ri* originated from the anaphoric marker *ri ... ri*, extending its functions to marking establishing, contextually unique, and bridging referents. The fact that a definite article like *ri* in Kaqchikel does not extend its use to deictic referents can be accounted for by the presence of other demonstrative markers in the grammar. In other words, deictic referents being expressed by demonstratives, there may simply be no functional pressure on the definite article to extend its use to deictic contexts.

Thus, it is very likely that many definite articles are compatible with deictic referents simply as a consequence of their origin as a demonstrative. If the source element of the definite article is

not compatible with deictic referents like the anaphoric marker *ri ... ri* in Kaqchikel, the resulting definite article is likely incompatible with deictic contexts as well.

5.1.4 Definite articles with absolutely unique referents

Absolutely unique referents are the only referents of their kind. Typical examples are the referents of ‘sun’ and ‘moon’. This makes the referents automatically identifiable, independently of the discourse situation. In some languages, for instance in English, the definite article is able to mark such referents even without any previous mention:

(220) **The moon** is bright tonight.

Absolutely unique referents may be expected to be compatible with the use of the definite article, given their inherent unambiguous identifiability. In Hawkins (1978: 115), for instance, such referents are grouped together with contextually unique referents, i.e. listed as contexts which require the definite article (in English). On the other hand, one may also expect that absolutely unique referents need not be marked by the definite article because they are already unambiguously identifiable and do not require a linguistic marker for that. In other words, one may ask what the function of the definite article is in such contexts. The aim of this section is to show that we find both types of behaviours in the languages of the world; in some languages, the definite article is used with absolutely unique referents like *the* in English, whereas other languages cannot use the definite article in such contexts. This means that the occurrence of a definite article with absolutely unique referents is a language-specific property; there is no general criterion for definite articles to be used with such referents.

Absolutely unique referents consist of only a small finite set of concepts. Usually, examples are restricted to ‘sun’ and ‘moon’, but, depending on cultural and ethnographical settings, we may find ‘earth’, ‘ocean’, ‘sea’, ‘rain’, ‘God’ and related concepts to be interpreted as absolutely unique. Also the expressions of tribes or cultures may belong to this set.

The crosslinguistic tendency of definite articles to mark absolutely unique referents or not is difficult to determine at this point, since this use is not necessarily discussed in the language descriptions. Nevertheless, from the cases in the sample that do mention the use of definite articles in such contexts, no trend in either direction is evident. The following examples illustrate both types of definite articles with respect to their compatibility with absolutely unique referents. Examples (221) to (224) show definite articles with absolutely unique referents in Mokpe (Bantu, Cameroon), Dime (South Omotic, Ethiopia), Lavukaleve (Papuan, Solomon Islands), and Apinayé (Nuclear-Macro-Je, Brazil). The definite articles of the latter three languages are shown

in (222a), (223a), and (224a) in a different definite context. Their compatibility with absolutely unique referents can be seen in (222b), (223b), and (224b), respectively.⁹

(221) ò ʒóŋgó ígbě / *ígbě.
 2SG NEG.look sun.ART:DEF / sun
 ‘Don’t look into the sun.’
 Mokpe (primary data)

(222) a. šiftay-ko šif-is ʔeen nítsob.
 shiftaye-GEN shoes-ART:DEF early childhood
 ‘The shoes of Shiftaye are from his childhood.’
 Dime (Seyoum 2008: 45)

b. ʔírf-ís múlmíl-ind.
 moon-ART:DEF round-F
 ‘The moon is round.’
 Dime (Seyoum 2008: 44)

(223) a. [Savata-m na] fin fafas hin.
 ninth-SG.M ART:DEF.M FOC:3SG.M fafas.M FOC:3SG.M
 ‘The ninth [fish] is a fafas.’
 Lavukaleve (Terrill 2003: 77)

b. [kui na] ia-re a-e-kiu-ge malav va suni
 sun.M ART:DEF.M be.hot-NONFIN O:3SG.M-SUB-die-ANT people.PL ART:DEF.PL all
 kini lagi-re lo-v.
 ACT shelter-NONFIN finish-PL
 ‘The sun was very hot, and the people all went and sheltered (in the shade).’
 Lavukaleve (Terrill 2003: 91)

(224) a. pa na pa iŋmã ate [pikap ja] nõr prãm ket.
 1.EMPH REAL 1 1.DAT 2.ERG earth ART:DEF lie.on.NONFIN want NEG
 ‘It is I who don’t want you to lie on **the ground**.’
 Apinayé (Cunha de Oliveira 2005: 86)

b. əw əbri ɲun wεʔε [čučüti ja] ně [čučüre ja] kət me pa
 yes then DS QUOT sun ART:DEF CONJ moon ART:DEF 3.ERG PL 1.ACC
 n-ipeč kačiw ate wa ri pa.
 RELAT-make PURP alone DU DEM live
 ‘Well, Sun and Moon, they lived [on Earth] by themselves, at the time they were to create us.’
 Apinayé (Cunha de Oliveira 2005: 304)

⁹See the online appendix for more examples of the definite articles in (222) to (223).

In other languages, the definite article does not occur with absolutely unique referents. This is shown in (225) to (228) for Bambassi (Blue Nile Mao, Ethiopia), Arawak (Arawakan, Suriname Guyana, French Guyana), and Ch’ol (Mayan, Mexico), and Biak (Austronesian, Indonesia). Again, examples (225a), (226a), (227a), and (228a) feature the definite articles of these languages in a different type of definite contexts. By contrast, we see in (225b), (226b), (227b), and (228b) that the definite articles are not used with absolutely unique referents. Note that in Biak, although the absolutely unique referent *swan* ‘sea’ is not marked by the definite article *anya* in (228b), it occurs with the inclusive-specific article *-ya*. That *swan* ‘sea’ is an absolutely unique referent is mentioned by van den Heuvel (2006: 219), who notes that the definite article is generally not required with “[c]oncepts that are generally known and therefore uniquely identifiable or ‘globally accessible’ [...]”

- (225) a. tí-f [íʃ ka:l-là] ha-tí-mí-á wi:-ín í-té íʃ-kol-là
 1SG-SBJ ART:DEF porridge-OBJ AFF-1SG-eat-DECL say-SS:NF 3SG-SBJ 3-PL-OBJ
 ha-mé:nt-á.
 AFF-tell-DECL
 ‘She told them, saying “I ate the porridge”.’
 Bambassi (Ahland 2012: 640)
- b. sáp-èt á:ns-íʃ ha-pòn-pòn-á.
 night-LOC moon-SBJ AFF-go.out-RED-DECL
 ‘The moon comes out at night.’
 Bambassi (Ahland 2012: 445)
- (226) a. De to Ebesilio ken na-balhosen-ka jahada de [to siko] khona.
 I be Ebesilio and their-leader-PERF here I ART:DEF.NHUM village at
 ‘My name is Ebesilio, and I am their leader here in the village.’
 Arawak (Pet 2011: 217)
- b. Ken hadali min-ka kho there-n hibir.
 and sun minimum-PERF not hot-SUB already
 ‘And the sun was already very hot.’
 Arawak (Pet 2011: 206)
- (227) a. tsa’=bi ochi tyi i-mali ch’èñ [li wiñik=i].
 PFV=REPORT enter PREP A:3-inside cave ART:DEF man=FIN
 ‘The man went into the cave.’
 Ch’ol (Vázquez Alvarez 2011: 246)
- b. che’ och-em=ix k’iñ, ...
 when enter-PART=already sun ...
 ‘When [the] sun is already down, ...’
 Ch’ol (Vázquez Alvarez 2011: 475)

- (228) a. Indya fyanu [rofan **anya**], ifnovku mankroder=i.
 so feed.3SG dog **ART:DEF** 3SG-with frog=ART:EXSPEC
 ‘So he took care of the dog, together with a **frog**.’
 Biak (van den Heuvel 2006: 204-205)
- b. Swan-ya i-vrin.
sea-ART:INSPEC 3SG-quiet
 ‘The sea was quiet.’
 Biak (van den Heuvel 2006: 219)

Although it may seem as if the use of the definite article with absolutely unique referents should be a later development of articles that have already extended their use to all the relevant contexts, this is not (necessarily) the case. For the emerging definite article in Old High German, Szczepaniak (2011: 74-77) and Flick (2020: 59), going back to Ouzoubar (1989), note that the use of the emerging definite article with absolutely unique referents such as *the Holy Spirit*, *the disciples of Jesus*, and *the Savior*, is in fact attested early on. For instance, in the Tatian translation (Otfrid) in the 9th century, human absolutely unique and contextually unique referents were already systematically marked by the definite article; (229) shows a few examples.

- (229) a. **ther** *heilant* ‘the (m.) Savior’
 b. **thie** *jungiron* ‘the (pl.) disciples of Jesus’ (lit. ‘youngers’)
 c. **thie** *Pharisei* ‘the (pl.) Pharisees’
 Old High German (Szczepaniak 2011: 74)

Szczepaniak (2011: 75) notes three other interesting patterns from the same Tatian translation. On the one hand, we see the first generic uses of human concepts that are marked by the emerging definite article. At the same time, the use of the article is not yet systematic with non-human pragmatic definite referents (e.g. anaphoric ones). This means that there is a large difference between human and non-human referents in terms of the applicability of the definite article. Nevertheless, inanimate absolutely unique referents, can already be observed to take the definite article, even if only sporadically. A few typical examples are given in (230).

- (230) a. **thiu** *sunna* ‘the (f.) sun’
 b. **ther** *mano* ‘the (m.) moon’
 c. **thiu/thaz** *worolt* ‘the (f./n.) world’
 Old High German (Szczepaniak 2011: 75)

Szczepaniak (2011: 75) shows for four inanimate absolutely unique referents how often they occur with or without the emerging definite article in Otfrid’s *Evangelienbuch*. For instance, for the noun *worolt* ‘world’, she counts 118 occurrences without and 33 with the definite article; the

numbers are 49 vs. 6 for the noun *himil* ‘heaven’, respectively (Szczepaniak 2011: 75). Such figures confirm that, indeed, these referents are only rarely marked by the article in the 9th century text, in contrast to animate absolutely unique referents, which according to Flick (2020), Ouzoubar (1989), Szczepaniak (2011) were already expressed using a definite article in a systematic way.

This means that the definite article in Old High German clearly did not extend its referential function in a “straight line” from most pragmatic to most semantic contexts, with absolutely unique referents at the very semantic end. Rather, the extension was subject to animacy effects (amongst other factors), developing first with human referents and then extending towards non-human and inanimate referents. However, considering only inanimate referents, the definite article did not necessarily first become systematically used with e.g. anaphoric or contextually unique referents before extending its use to absolutely unique referents. Szczepaniak (2011) and Flick (2020) argue that the emerging definite article is used with such inanimate absolutely unique referents for discourse-pragmatic reasons, and not to indicate their uniqueness (and therefore, its unambiguous identifiability). Thus, the emerging definite article is used in order to draw the reader’s attention to the referent and mark it as relevant for the following discourse. Going into further details of the development of the definite article in Old High German would go beyond the scope of the present study. The important point of this digression was to show that the compatibility of definite articles with absolutely unique referents does not necessarily correspond to a later stage of the development of the definite article, as it would be tempting to assume based on the scales from most pragmatic to most semantic types of definite referents discussed in Section 3.4.

To sum up, this section showed that definite articles in the languages of the world can but do not have to be used with absolutely unique referents. And if they do, one should be careful to explain this as a later functional extension in the development of the definite article or to expect that definite articles may necessarily reach this state at some point.

5.1.5 Definite articles with proper nouns

Similarly to the compatibility of definite articles with absolutely unique referents, their occurrence with proper nouns (including names) varies substantially across languages.

Again, we may expect definite articles to occur with proper nouns and names because their referents are usually unambiguously identifiable. On the other hand, one could argue that such referents are inherently definite and may not need additional marking as such by a definite article. Again, we find examples of both behaviors across languages, with no evident trend either. There are four main co-occurrence patterns of definite articles and names that can be distinguished. The article can be systematically required with names, its use can be blocked, or the article can be optional. In addition, some language use a different marker with names and proper nouns in contrast to the definite article used with common nouns. Besides these four main patterns, the

use of the definite article with proper nouns may depend on additional modifiers. In English, for instance, the definite article is not used with names like *(*the) Julie*. If restrictively modified, however, the use of the article becomes obligatory, as in **(the) Julie I used to know* (cf. [Matushansky 2006](#)).¹⁰ Due to the scope of the present study, I will only focus on the major patterns and provide a general overview of the crosslinguistic variation regarding the use of definite articles with proper nouns and names.

The following examples illustrate the co-occurrence of the definite article with names or proper nouns in Arawak (Arawakan, Suriname Guyana, French Guyana), Irish (Indo-European, Ireland, UK), and Tarahumara (Uto-Aztecan, Mexico). The use of the definite article with (certain types of) proper nouns is generally required in these languages.¹¹ The examples in (231a), (232a), and (233a) illustrate the use of the definite articles with a common noun; and we see them together with a proper noun in (231b), (232b), and (233b).

- (231) a. **li** **falhetho**
 ART:DEF white.man
 ‘the white man’
- b. **li** **Wim**
 ART:DEF Bill
 ‘Bill’
- Arawak ([Pet 2011: 52-53](#))

- (232) a. **an** **bád**
 ART:DEF boat
 ‘the boat’
- b. **an** **Cháisc**
 ART:DEF Easter
 ‘Easter’
- Irish ([Stenson 2008: 35-36](#))

- (233) a. **echi** **torí**
 ART:DEF rooster
 ‘the rooster’
- b. **echi** **Antonio Loera**
 ART:DEF Antonio Loera
 ‘Antonio Loera’
- Tarahumara ([Cohen 1998: 70,129](#))

In other languages, definite articles can combine with proper nouns and names, but their co-occurrence is optional or subject to inter-speaker variation, as in German, shown in (234).

¹⁰I thank Klaus von Heusinger for bringing this to my attention.

¹¹Cf. [Pet \(2011: 86\)](#) for Arawak, [Stenson \(2008: 36\)](#) for Irish, and [Cohen \(1998: 144\)](#) for Tarahumara.

- (234) (der) Thomas
 ART:DEF.M Thomas
 ‘Thomas’
 German (primary data)

The definite articles may be also be absent with proper nouns. This is what we observe in, for instance, Sheko (Dizoid, Ethiopia) or Biak (Austronesian, Indonesia).¹² In the following two examples, we see the use of the definite article with a common noun in (235a) and (236a), and opposed to that its absence with proper nouns (names) in (235b) and (236b).

- (235) a. ts’ahāfi-ñ-s
 clerk-ART:DEF-M
 Sheko (Hellenthal 2010: 144)
- b. gaaa-kñ āšū fan=á-k-ə.
 Gaana-DAT leg be.broken=3SG.M-REAL-STI
 ‘Gaana’s leg is broken.’
 Sheko (Hellenthal 2010: 238)
- (236) a. [mankroder anya] ma [romá-mkun anya] d-úf i.
 frog ART:DEF TOP child-little ART:DEF 3SG-pick.up 3SG
 ‘As for the frog, the little child picked it up.’
 (van den Heuvel 2006: 204-205)
- b. Sahulata ima rya ma ...
 Sahulata 3SG.FOC go.3SG hither ...
 ‘Sahulata, she came here ...’
 Biak (van den Heuvel 2006: 216)

The last type to be mentioned here corresponds to languages that use a marker different from the definite (or specific, or referential) article with proper nouns. This is often found in Austronesian languages. Example (237) illustrates this for the referential article in Rapa Nui (Austronesian, Chile). Here, the article *te* occurs with common nouns, while proper nouns need to be marked by *a*.

- (237) he oho a Hotu ki te hare.
 NTR go ART:REF.PROP Hotu to ART:REF house
 ‘Hotu went home.’
 Rapa Nui (Kieviet 2017a: 102)

Note that I do not treat markers like *a* in Rapanui as articles (even though they are typically referred to as proper or personal articles). As was mentioned in Chapter 2, these markers do not have referential functions but mark lexical properties of the referents.

¹²Cf. Hellenthal (2010: 142) for Sheko and van den Heuvel (2006: 221) for Biak.

5.2 Anaphoric articles

Anaphoric articles were defined in Section 3.3.1 as articles that systematically occur with anaphoric referents, but which cannot be used to encode deictic or contextually unique referents. They correspond to what was introduced as a strong definite article in Schwarz (2009) and discussed in a number of other studies (Arkoh & Matthewson 2013, Ingason 2016, Irani 2019, Jenks 2018, Ortman 2014, Schwarz 2013, 2019, Šereikaitė 2019). In this tradition, strong definite articles are contrasted with weak definite articles, which cannot be used in anaphoric or other pragmatic definite contexts and are only used to mark semantic definites (cf. Section 5.3). In this study, I use the label “anaphoric” instead of strong definite article for two reasons. First, the main function of such articles is to mark anaphoric referents. Second, the label strong suggests that the article is generally used with different pragmatic definite referent types, i.e. deictic, situationally unique, recognitional, establishing referents. However, this is not the case. Such articles are not able to mark deictic referents (they may be compatible with them, though), and they vary in their ability to occur with recognitional, establishing, and situationally unique referents (the latter of which is usually not encoded by anaphoric articles). Thus, to avoid confusion, I call such articles anaphoric articles to emphasize that the coding of anaphoric articles is their main function. In this section, I present the three examples of anaphoric articles from Limbum (Section 5.2.1), Komnzo (Section 5.2.2), and Akan (Section 5.2.3). In the last part, Section 5.2.4, I briefly describe the use of anaphoric articles in establishing and recognitional contexts.

5.2.1 Limbum

Limbum is a Grassfields Bantu languages spoken in the Donga-Mantung department in the North West region of Cameroon. It has approximately 130 000 speakers.¹³ Like other Grassfields Bantu languages, Limbum has a complex tonal system and is isolating in terms of morphology. It does not have any other articles than the anaphoric article *fō* presented in this section.

In order to show that the marker *fō* is an anaphoric article, we will first look at the demonstrative system. According to Fransen (1995: 146), Limbum features a tripartite system of spatial demonstrative markers: one value expresses closeness to the speaker, the second one closeness to the hearer, and the third value marks distance from both the speaker and the hearer. The forms are shown in Table 5.5.

¹³Cf. <https://www.ethnologue.com/language/lmp>.

Table 5.5: Limbum demonstratives

closeness S	<i>šá (čá)</i> ¹⁴
closeness H	<i>ánà</i>
distance S & H	<i>šà (čà)</i>

In addition to these spatial demonstratives mentioned in Fransen (1995), Limbum has another marker *fō* which qualifies as an anaphoric article. As will be shown in this section, *fō* systematically occurs in anaphoric contexts, but it cannot be used to mark other types of definite referents such as deictic and contextually unique referents.

That the marker *fō* is obligatory in anaphoric contexts is illustrated in examples (238) and (239). In (238b), we see that the referent of *ndzínwě* ‘woman’ uttered by speaker B refers back to the same referent uttered by speaker A in (238a). The anaphoric referent in B’s answer has to be marked by *fō*. Example (238c) shows that the use of the demonstrative *ánà*, on the other hand, is infelicitous in B’s answer.

- (238) a. *mú mfè? mē mū bā?, mē mū yé ndzínwě bá bōō báā ó tʃútèh mí*
 when time I PST2 arrive I PST2 see **woman** and children two 3PL sit in
kísìn.
 kitchen
 ‘When I came home, I found a woman with two children sitting in my kitchen.’
- b. *wè mū rīŋ [ndzínwě *(fō)] ā?*
 2SG PST2 know woman ART:ANA Q
 ‘Did you know the woman?’
- c. *wè mū rīŋ [ndzínwě *ánà] ā?*
 2SG PST2 know woman DEM Q
 intended: ‘Did you know the woman?’
 Limbum (primary data)

The same pattern is shown for the anaphoric referent of *mú* ‘child’ in (239), introduced in (239a). In (239b), *mú fō* refers back to the previously mentioned referent; *fō* is obligatory in this context. Example (239c) shows that using *ánà* instead of *fō* is less acceptable (even though not as infelicitous as its use in (238c)).

- (239) a. *àmbò djò? à m cī rò bzhī ā tárté ye o mú mò? ...*
 when elephant s.3SG PST3 PROG search food 3SG stumble 3SG with **child** one ...
 ‘When it was looking for food, the elephant stumbled into a child ...’

¹⁴The versions of the demonstratives given in brackets correspond to the forms used in the variety of my consultant.

- b. ... [mú *(fō)] à kē? á cí wārī.
 ... child ART:ANA 3SG start INF PROG cry
 ‘... The child started to cry.’
- c. ... [mú ?ánà] à kē? á cí wārī.
 ... child DEM 3SG start INF PROG cry
 ‘... The child started to cry.’
 Limbum (primary data)

A similar context is given in example (240). Here, speaker A expresses that she is thirsty and asks for water in (240a). Speaker B confirms that there is water in the kitchen. Speaker A goes to the kitchen, where she utters (240b). The referent of *mdzīp* in (240b) requires to be marked by *fō*, even though it is not a typical anaphoric referent because the antecedent in (240a) is nonspecific. In addition, (240c) shows that the distal demonstrative *ánà* is not felicitous in this context.

- (240) a. ndōŋ à yúfī mè. à cōr **mdzīp** mō? ā?
 neck is dry 1SG you have water some Q
 ‘I am thirsty. Do you have (some) water?’
- b. vù mū càr? ì cī yē [mdzīp *(fō)] kā?
 come DIM small I PROG see water ART:ANA NEG
 ‘Can you come? I do not see the water.’
- c. vù mū càr? ì cī yē [mdzīp *ánà] kā?
 come DIM small I PROG see water DEM NEG
 Limbum (primary data)

Although *fō* is systematically used to mark anaphoric referents, it is not obligatory in all anaphoric contexts. Example (241) below shows that with abstract referents, for which a generic reading is available in addition to the situationally grounded one, the use of the anaphoric article is not obligatory. The “optionality” of the anaphoric article in this case is most probably due to the additional generic reading, but it also seems to involve a stylistic contrast. In fact, the English translation of (241) shows the same effect for the definite article in English. Therefore, I assume that *fō* can nevertheless be considered an anaphoric article, allowing for exceptions due to style or other additional factors.

- (241) ŋkār à bbītī rkòŋnì [rkòŋnì (fō) / (*ánà)] à bītī rbàni.
 friendship has become love love ART:ANA / DEM has become hatred
 ‘Friendship became love. (The) love became hatred.’
 Limbum (primary data)

In addition, we find some anaphoric contexts in which the anaphoric article *fō* and the demonstrative *ánà* can both be used and are competing against each other. Example (242) below shows that in certain cases, both the anaphoric article *fō* as well as the spatial deictic *ánà* can be used:

- (242) a. $\eta w\grave{e}$ m̀ò à m cī cū mi Nkambe.
 man one 3SG PST3 PROG live in Nkambe
 ‘There used to live a man in Nkambe.’
- b. [$\eta w\grave{e}$ f̄o / áná] à m tʃur bō sàmbá.
 man ART:ANA / DEM 3SG PST3 have children seven
 ‘The man had seven children.’
- Limbum (primary data)

Such a situation is similar to other languages with a definite article that “competes” with a stronger demonstrative marker in certain contexts. The demonstrative may be used for reasons of emphasis, due to the discourse prominence of the reference, or the distance between the antecedent and the anaphoric referent. The same can be assumed to hold for the distribution between a spatial demonstrative like *ánà* and an anaphoric article like *f̄o* in Limbum. A more detailed analysis of the factors that condition the choice between those two markers in those contexts in which both are felicitous would go beyond the scope of the present study. The important point is that finding this kind of variation does not exclude Limbum *f̄o* from being analysed as an anaphoric article. The variation rather corresponds to what we find between other types of articles and demonstratives (or pronouns in general) in other languages as well.

In addition to the previous examples showing that *f̄o* is systematically used to mark anaphoric referents, it can be shown that the use of *f̄o* is not felicitous with other types of definite referents. Example (243) shows that the anaphoric article cannot be used with a deictic referent; a demonstrative, e.g. the distal *čà*, is used instead. If the speaker neither uses a demonstrative nor points towards the dog present in the discourse situation, *f̄o* can nevertheless not mark the referent of *\eta gw\grave{e}* ‘dog’ as situationally unique, and a bare noun is used instead.

- (243) í mū nyan [$\eta gw\grave{e}$ (čà)].
 3SG PST2 feed dog DEM:DIST
 ‘He fed that dog.’
- Limbum (primary data)

Example (244) shows an instance of rel-bridging. The bridging referent, expressed by *\eta w\grave{e} rsā\eta* ‘author’, is identifiable due to its link to the referent of *\eta wā?* ‘book’. Nevertheless, *f̄o* cannot be used to mark the referent of *\eta w\grave{e} rsā\eta* ‘author’ as definite. This example shows that rel-bridging is not necessarily marked in the same way as anaphoric referents and should therefore be treated as a separate (yet similar) referential function.

- (244) Tāntōh ā mū yū $\eta wā?$ mō; [$\eta w\grave{e}$ rsā\eta (*f̄o)] á $\eta w\grave{e}$ -mbòm.
 Tantoh has PST2 buy book one person writing ART:ANA is person-Wimbum
 ‘Tantoh has bought a book; the author is a Wimbum person.’
- Limbum (primary data)

The same effect can be observed for contextually unique referents. The two examples (245) and (246) below illustrate that the anaphoric article *fō* cannot be used to mark this referential function.

(245) [ntāh (*fō)] yō fē (à là? nā)?
 market.CL9 ART:ANA be.CL9 where in village this?
 ‘Where is the market (in this village)?’
 Limbum (primary data)

(246) [tō ntā? (*fō)] tvír mfè? àmbò bbìr kā?
 head chair.AM ART:ANA have time about wars NEG
 ‘The President does not care (does not have time for) about the strikes.’
 Limbum (primary data)

Recognitional referents appear to be marked by either the demonstrative *čà* or by the anaphoric article *fō*. The use of the demonstrative *čà* in such contexts is shown in examples (247) and (248), where the use of *fō* is infelicitous according to my consultant. In both examples the referents of *ndíŋwè* and *mbví* are mentioned for the first time in the current discourse situation.

(247) mè gí ā cí dē? wèr [ndíŋwè čà / #fō].
 I PST1 1SG PROG talk with woman DEM:DIST / ART:ANA
 [Context: A asked B to talk to a certain woman; B then meets A telling her that she did talk to the woman.]
 ‘I just talked to that woman.’
 Limbum (primary data)

(248) wà? à kūcī [mbví čà / #fō] ā?
 POL you remember goat DEM:DIST / ART:ANA Q
 ‘Do you remember that goat (we used to have)?’
 Limbum (primary data)

However, example (249) shows a context in which *fō* is used with a recognitional referent. Here, the referent of *bìr* ‘war/strike’ is mentioned for the first time in the current discourse, which is situated in the anglophone part of Cameroon in 2017, where strikes against the governmental repression correspond to knowledge and experience shared between the speaker and the hearer. In this case, *bìr* ‘war/strike’ can be marked by the anaphoric article.

(249) n̄sūŋ m̄5? yō àmbò [bìr (fō)] ā?
 news any COP about war ART:ANA Q
 ‘Are there any news about the strike?’
 Limbum (primary data)

The anaphoric article is also used in establishing contexts. Examples (250) and (251) show two such contexts, where a referent is established as identifiable by *fō* and then made identifiable by a following relative clause.

(250) [mu fō] zhì i mu do à rìŋ bi gwè yaŋ.
 child ART:ANA COMP 3SG PST2 go to forest NEARFUT fall sick
 ‘The child that went to the forest will fall ill.’
 Limbum (primary data)

(251) vʌ nì [msaŋ fō] zhì sèè àmu yuh nìŋkòr.
 come with rice ART:ANA COMP 1PL PST2 buy yesterday
 ‘Bring the rice that we bought yesterday.’
 Limbum (primary data)

Table 5.6: The distribution of *fō* in Limbum

referent types	use of <i>fō</i>
R _{deictic}	✗ (243)
R _{anaph}	✓ (238), (239), (240), (241), (242)
R _{recog}	(✓) (247), (248), (249)
R _{establ}	✓ (250), (251)
R _{rel-bridge}	✗ (244)
R _{sit.unique}	✗ (243)
R _{cont.unique}	✗ (245), (246)

To conclude this section about *fō* in Limbum, we saw that the marker is systematically used in anaphoric and establishing contexts. Besides the demonstrative, the anaphoric article is also able to encode a recognitional referent. Anaphoric *fō* in Limbum was also shown to be infelicitous in expressing other definite functions, most importantly contextually unique and u-bridging. The referential functions of the anaphoric article *fō* in Limbum are summarized in Table 5.6.

5.2.2 Komnzo

Komnzo, a Papuan language of the Yam group, is spoken in the Southern part of New Guinea (Papua New Guinea). Döhler (2018) estimates that the language has between 150 and 250 speakers. Besides the anaphoric article *ane*, Komnzo has what I call a presentational article. The use of *ane* with anaphoric referents is shown in examples (252) and (253). It is an anaphoric article and not a demonstrative or a definite article, because it systematically occurs with anaphoric but with neither deictic, nor contextually unique referents. Döhler (2018: 110) describes the marker in the

following way: “[...] *ane* has no spatial reference, but it is used for anaphoric reference. It marks a referent which has been established in the preceding context.”

(252) a. **wämne** ...yf füni yé firra=n bä
 tree ...name füni be.S:3SG.M.NPST.IPFV firra=LOC DEM:MED
 ykogr.
 stand.S:3SG.M.NPST.STAT

‘The name of the tree is füni. It stands there in Firra.’

b. nä kayé fthé boba gnyako nima kwa
 ART:PRES yesterday when DEM:MED.ALL go.S:3SG.IMP like.this FUT
 ymarwr ... [**ane** kafar wämne].
 see.S:2SG.O:3SG.M.IPFV ... ART:ANA big tree

‘When you go there some day, you will see it ... that big tree.’

Komnzo (Döhler 2018: 367)

(253) a. **bthan kabe** fthé fenz yona-si bänemr zrethkäfth
 magic man when body.liquid drink-NMLZ RECOG.PURP start.IRR.PFV.S:3PL
 mätrak-si=r.
 take.out-NMLZ=PURP

‘When **the sorcerers** drink the body fluids, they start by bringing out this one.’

b. ... fthé fof krefar [**ane** bthan kabe] bobo fokam zn=fo
 ... when EMPH set.off.IRR.PFV.S:3SG ART:ANA magic man MED.ALL grave place=LOC
 fokam mnz=fo sikwankwan=me zbär thd.
 grave house=LOC secret=INSTR night middle

‘... **the sorcerer** sets off to go to the grave yard, to the grave house. He goes secretly in the middle of the night.’

Komnzo (Döhler 2018: 408-409)

While the anaphoric article *ane* does not mark a spatial deictic referent as identifiable as such, it is compatible with deictic referents that are marked by another demonstrative identifier, as in example (254).

(254) fintäth **ane** z=yé yem=aneme dagon.
 fintäth ART:ANA DEM:PROX=be.3SG.NPST cassowary=POSS:NSG food

‘This fintäth (fruit) here is the cassowaries’ food.’

Komnzo (Döhler 2018: 111)

Importantly, *ane* does not occur with contextually unique referents to mark them as identifiable. In example (255), the referent of *ɲars* ‘river’ is unambiguously identifiable for all discourse participants because it refers to a contextually salient river, which has not been mentioned yet in the preceding discourse segments. The referent of *dödö* ‘broom’ in (256) is situationally unique, given

that cleaning involves a single broom by default. Again, we see that the anaphoric article *ane* is not used with the noun.

- (255) kabe matak erä nima z bramöwä kwafarkwrth nima
 man nothing be.NPST:IPFV.3PL like.this already all set.off.PST:IPFV.3PL like.this
 erä **ɲars=fo** ...
 be.NPST:IPFV.3PL **river=ALL** ...
 ‘Nobody is here. All the people left this way to the river ...’
 Komnzo (Döhler 2018: 365)

- (256) **dödö** thfefaf ane zurenwrmo mnz fath thwafiyokwrm.
broom hold.ITER.3SG ANA sweep.PST:DUR.3SG.F house clear.place make.DUR.S:SG.O:3PL
 ‘She always grabbed the broom, swept the house and cleaned it for them.’
 Komnzo (Döhler 2018: 362)

As for recognitional referents expressed by a noun, it is not clear whether *ane* can be used in such a context. Komnzo has a separate recognitional marker *baf*, which is used only pronominally (Döhler 2018: 112). Therefore, I do not consider it as a recognitional article here, although the contexts that *baf* ‘RECOG’ appears in closely mirror the ones in which we find recognitional articles, involving the typical prosodic break and often functioning as a filler. Two examples are given in (257) and (258).

- (257) wati nzedbo zanrifhath mayawa=medbo rouku **bäne=fo** (.) masu=fo.
 then 1NSG.ALL send mayawa=ALL.ANIM.NSG rouku **RECOG=ALL** (.) masu=ALL
 ‘Then they sent to world to us ... to the Mayawas in Rouku ... to there ... to Masu.’
 Komnzo (Döhler 2018: 150)

- (258) fi fenz ane **bänemr=nzo** rä (.) tmä yari-si=r.
 but body.liquid ART:ANA **RECOG.PURP=only** be (.) strength give-NMLZ-PURP
 ‘but the body liquid is only for this ... for giving power.’
 Komnzo (Döhler 2018: 157)

For establishing referents, on the other hand, we find ample evidence in Döhler (2018) that *ane* is used to mark them as identifiable. Examples (259) and (260) illustrate the use of the anaphoric article *ane* in establishing contexts, occurring with referents that have not been mentioned in the previous discourse segments.

- (259) nagayé=aneme znsä=n zwäfonz [ane
 children=POSS.NSG work=LOC be.caught.by.nightfall.RECPST:PFV.1SG **ART:ANA**
 gathagatha=me k-kauna mane egathikwroth].
 bad=INSTR RED-thing which leave.NPST:IPFV.3PL
 ‘I was caught by nightfall while working for the children sorting those things which they leave scattered around.’
 Komnzo (Döhler 2018: 368)

- (260) waisamen ... waisamen [ane kar yf rä mobo
 waisamen ... waisamen ART:ANA place name be.NPST:IPFV.3SG.F where.ALL
 zwamnzm] ... mrab fr thd=en.
 dwell.PST:DUR.3SG.F ... bamboo grove middle=LOC
 ‘Waisamen ... Waisamen is the name of that place where she was living ... in the middle of
 a bamboo grove.’
 Komnzo (Döhler 2018: 371)

Thus, the main function of *ane* in Komnzo was shown to be the marking of anaphoric referents as unambiguously identifiable. At the same time, we saw that it cannot be used to mark contextually unique or deictic referents, although it seems to be compatible with the latter function. Because of this distribution, *ane* is an anaphoric article. Its absence with recognitional referents, otherwise expected as a default extension of anaphoric markers, could be attributed to the availability of a separate marker that fills this function (although not fully adnominally). Like the anaphoric article in Limbum presented in the previous section, we saw that Komnzo *ane* is used to establish referents. The referential functions of the anaphoric article *ane* are summarized in Table 5.7.

Table 5.7: The distribution of *ane* in Komnzo

referential functions	use of <i>ane</i>
R _{deictic}	(✓) (254)
R _{anaph}	✓ (252), (253)
R _{recog}	✗ (257), (258)
R _{establ}	✓ (259), (260)
R _{sit.unique}	✗ (256)
R _{cont.unique}	✗ (255)

5.2.3 Akan

Another anaphoric article can be found in Akan, a Kwa language (or group of mutually intelligible dialectal varieties) spoken in Ghana. Akan is one of Ghana’s main languages and is estimated to have over 8 million speakers.¹⁵ In addition to the marker *nó* that I analyse as an anaphoric article, Akan has the exclusive-specific article *bi*, described in Section section 6.3.2.

There are a number of previous studies that have dealt with *nó* in Akan, e.g. Amfo (2010), Arkoh (2011), Arkoh & Matthewson (2013), Bombi (2018), Bombi et al. (2019), Fretheim & Amfo (2008), Korsah (2017), Saah (1994).¹⁶ In earlier work, *nó* is referred to as definite marker or definite

¹⁵Cf. <https://www.ethnologue.com/language/aka>.

¹⁶I use the spelling of *no* here following the Standard orthography of Akan. In some examples cited from the literature, the anaphoric article is spelled as *nó* which represents the IPA transcription of the vowel and the high tone. Besides their spelling, those two forms are identical.

determiner (e.g. Amfo 2010, Arkoh 2011, Saah 1994, 2010); Fretheim & Amfo (2008) analyse *no* as both a definite article and a demonstrative.

Inspired by the proposal of weak and strong definite articles, Arkoh & Matthewson (2013) argue that *nó* is a (strong) familiar definite article. They offer a detailed discussion of *nó* compared to the strong definite article in German, presenting different types of definite contexts that *nó* is and is not used in. Arkoh & Matthewson (2013) conclude that it indeed only occurs with pragmatic (strong) definite referents, whereas it cannot be used with semantic (weak) definites. Since then, Akan *nó* has been one of the showcase examples of strong definite articles in the literature (e.g. Jenks 2018, Schwarz 2013, 2019). However, other scholars have recently argued that Akan *nó* is in fact a definite article rather than an anaphoric article (Bombi 2018, Bombi et al. 2019).

To relate the ongoing discussion of Akan *nó* to the crosslinguistic patterns of articles analysed in present study, this section deals with Akan *nó*, mainly summarizing and commenting on the relevant patterns from the literature, including a few additional examples. My analysis of *nó* falls somewhat in between the accounts of Arkoh & Matthewson (2013) and Bombi (2018). I analyse *nó* as an anaphoric (i.e. strong definite) article, even though *nó* appears to have extended its functions towards semantic definite uses, which admittedly makes it a less prototypical anaphoric article than the ones in Limbum or Komnzo presented in the previous sections. The reason for treating it as an anaphoric article nevertheless is that *nó* does not occur with contextually unique referents. Therefore, I conclude that it does not, at least at this point, cover the whole range of referential functions that a definite article is assumed to cover.

Example (261) shows the article *nó* in an anaphoric context: after the referent of *èkùtú* ‘orange’ is introduced, it has to be mentioned together with the article in (261b). The same can be seen in examples (262) and (263), where *nó* is used to mark the referents of *àtààdéé* ‘dress’ and *àbrántéé* ‘boy’ as anaphoric as well.

- (261) a. Mò-tó-ò èkùtú bí.
 1SG-buy-PST **orange** ART:EXSPEC
 ‘I bought an orange.’
- b. [Èkùtú nó] yè dèw déè.
 orange ART:ANA be nice so
 ‘**The orange** is/was so nice.’
 Akan (Amfo 2010: 52-53)
- (262) a. Mè-tò-ò àtààdéé bí ònóra.
 1SG-buy-PST **dress** ART:EXSPEC yesterday
 ‘I bought a dress yesterday.’

- b. [Àtààdéé **nó**] yé fè.
 dress ART:ANA COP nice
 ‘The dress is nice.’
 Akan (Bombi 2018: 148)

- (263) a. Kofi hù-ù mààmé né àbrántéé.
 Kofi see-PST woman COORD young.man
 ‘Kofi saw a woman and a boy.’
 b. [Àbrántéé **nó**] kyèá-à Kofi.
 young.man ART:ANA greet-PST Kofi
 ‘The boy greeted Kofi.’
 Akan (Bombi 2018: 148)

As was mentioned above, Arkoh & Matthewson (2013) argue for familiarity as a condition for the use of *nó*. They show that the referent has to be previously mentioned within a given discourse situation in order to be marked with the article; its use is infelicitous otherwise. Arkoh & Matthewson (2013) provide the real-life example shown in (264), where the first mention of a referent with *nó* is infelicitous and reference cannot be established by the hearer. Speaker A, having come home from university, utters (264a). The hearer of (264a) is confused and utters (264b), asking for clarification from speaker A about the identity of the referent of *àbòfrá* ‘child’. Thus, using *nó* is not successful for marking the referent as identifiable in this context.

- (264) a. [Àbòfrá **nó**] bá-à há.
 child ART:ANA come-PST here
 ‘The child came here.’
 b. Íbèn àbòfrá à? À-n-ká àbòfrá bíará hò ásém ò!
 which child Q PST-NEG-say child every self case EXPL
 ‘Which child? You did not say anything about any child oh!’
 Akan (Arkoh & Matthewson 2013: 8)

Arkoh & Matthewson (2013) also show that *nó* cannot occur with a (spatial) deictic referent if the referent is mentioned for the first time in the discourse situation. An example is given in (265). Here, the context is an out of the blue mention of the dog with no deictic gesture. The dog is visible when (265a) is uttered. In (265a), speaker A attempts to mark the referent of *bòdóm* ‘dog’ as situationally unique using *no*. That this is not felicitous can be seen from speaker B’s question in (265b), asking about the identity of the referent mentioned. Therefore, it is clear that speaker B could not construct the intended reference to the deictic and situationally unique referent.

- (265) a. Kwèsí bó-ò [bòdóm **nó**].
 Kwesi beat-PST dog ART:ANA
 ‘Kwesi beat the dog.’

- b. Íbén bòdóm á?
 which dog Q
 ‘Which dog?’

Akan (Arkoh & Matthewson 2013: 12)

However, Arkoh & Matthewson (2013) mention that (265a) is felicitous in the same context if the utterance is accompanied by a pointing gesture. This suggests that once deictic reference is established by the gesture, *nó* can mark the referent as situationally unique. Similarly, Bombi (2018) shows that the anaphoric article *nó* can be used together with a demonstrative, the latter of which marks the referent as a deictic referent. This provides further evidence for *nó* being compatible with deictic reference in cases in which it marks the referent as situationally unique. An example is given in (266).

- (266) Mè-pè [sàá àtààdéé **nó**].
 1SG-like DEM dress ART:ANA
 [Context: Kofi and Amma are in the market. Amma disappears and comes back with one dress / several dresses in her hands. Kofi says:]
 ‘I like that dress [pointing at Amma’s dress].’
 Akan (Bombi 2018: 151-152)

Another pragmatic definite function is the recognitional one. A recognitional referent is unambiguously identifiable even at first mention because of shared (personal) knowledge between the speaker and the hearer. Anaphoric articles may be able to occur in such contexts as well, probably by functional extension (cf. Section 4.1). This is what Arkoh & Matthewson (2013: 17) note: “[...] Akan *nó* does not enforce anaphoricity, but instead can be used even for referents mentioned in the remote past, for instance a year ago”. An indication for *no* being used with recognitional referents is given by Saah (1994: 152). Consider example (267).

- (267) [Mbofra **nó**] wɔ [dan **nó**] mu.
 children ART:ANA be room ART:ANA in
 ‘The children are in the room.’
 Akan (Saah 1994: 152)

Referring to example (267), Saah (1994: 153) remarks that “[t]he sentence can be paraphrased as: *The children (You and I know / have talked about) are in the room (that you and I know / have talked about)*”. This additional explanation suggests that the referents of *mbofra* ‘children’ and *dan* ‘room’ were not necessarily mentioned before in the current discourse situation, and that *nó* can mark them as recognitional.

The other function that anaphoric articles are expected is to mark establishing referents. Those are referents which are marked as identifiable, while the information that is necessary for their

unambiguous identification by the hearer is only provided in immediately following discourse segment, often by restrictive relative clauses. In Akan, the anaphoric article *nó* can be used in such contexts, even though its use does not appear to be systematically required. Examples (268) and (269) show two examples of establishing referents which are marked by *nó* and whose unambiguous identifiability is established in the following relative clause. Describing those two examples, Arkoh (2011: 76) notes that “[...] the familiarity of the nouns can be forced in special contexts where there is detailed description of the referents”. Thus, in such contexts, the referents of *ásùpítì* ‘hospital’ and *sikàkòràbíá* ‘bank’ do not have to be previously mentioned in order to occur with *nó*, marking them as identifiable on the basis of the information provided in the relative clause.

- (268) Mé-é-hù [ásùpítì *nó*] àà Kwámì kó-ɪ *nó*.
 SBJ:1SG-PERF-see hospital ART:DEF REL Kwame go-PST CD
 ‘I have seen the hospital Kwame visited.’
 Akan (Arkoh 2011: 75)

- (269) [Sikàkòràbíá *nó*] àà miyî siká wò *nó* á-hyìw.
 bank ART:ANA REL take.PST.1SG money at CD PERF-burn
 ‘The bank which I took money from is burnt.’
 Akan (Arkoh 2011: 76)

In addition to the use of the anaphoric article *nó* together with nouns to express an establishing referent, *nó* also appears as a clausal determiner (glossed CD) in examples (268) and (269). I will not discuss the use of *nó* as a clausal determiner here; suffice it to say that its development is certainly diachronically related to the development of *nó* as an anaphoric article.¹⁷

Returning to the use of the anaphoric article *nó* with establishing referents, other examples similar to (268) and (269) above show that *nó* is not used or optional in establishing contexts. Consider (270), where the anaphoric article *nó* can but does not have to be used, with no apparent change in meaning in its presence or absence according to Saah (2010: 5).

- (270) [Abofrá (*nó*)] àà ɔ-kó-ɔ hó *nó* bé-yare.
 child ART:ANA REL 3SG-go-PST there CD FUT-be.sick
 ‘The child who went there will fall ill.’
 Akan (Saah 2010: 95)

Finally, Arkoh & Matthewson (2013) give the example shown in (271) to argue that the use of *nó* is not felicitous with an establishing referent. They introduce (271) as follows: “[I]f the hearer has

¹⁷For more details on the clausal determiner in Akan and other (mostly) West African languages, see Aboh (2005), Ameka (1991), Bombi et al. (2019), Fretheim & Amfo (2008), Grubic & Renans (2016), Korsah (2017), Lefebvre (1992), Saah (1994, 2010), and references therein.

no knowledge of the referent of the noun phrase the sentence will be odd. For instance, consider a context where Esi visits her friend Ama and in conversation, Ama utters (271). Suppose that Esi has no prior knowledge of the said cassava.”

- (271) ?? É̀sì fá [bà̀nkyí nɔ́] áà ó-gú kè̀ntsén mù nɔ́ brà.
 Esi take cassava ART:ANA REL it-pour basket in CD come
 ‘Esi, bring the cassava that is in the basket.’
 Akan (Arkoh & Matthewson 2013: 9)

Given such examples of establishing referents, the distribution of the anaphoric article *nó* in such contexts seems far from clear. It may be the case that the use of *nó* is being extended in such contexts but that it does not (yet) occur in a systematic way. It may also be the case that many additional factors, e.g. the lexical properties of the noun, the properties of the predicates used, etc. may complicate the picture. A more detailed analysis of such contexts lies outside of the scope of this study; importantly, the examples clearly show that *nó* is able to mark (certain types of) establishing referents.

Akan *nó* is used in rel-bridging contexts as well, as is shown by the following examples (272) and (273). In (272b), we see that the anaphoric article can be used with the bridging referent of *ɔ̀twerefoɔ* ‘author’. However, as (272c) shows, using a possessive marker is an alternative strategy. In that case, the article cannot be used. In example (273), we see bridging between the event of ‘arriving’ that the referent of *baase* ‘bus’ can refer back to. The anaphoric article has to be used with the noun *baase*.

- (272) a. Kofi tɔ̀-ɔ̀ krataa bi ...
 Kofi buy-PST book ART:EXSPEC ...
 ‘Kofi bought a book ...’
 b. ... [ɔ̀twerefoɔ *(nɔ́)] fi Kumase.
 ... writer ART:ANA come.from Kumasi
 ‘... The author is from Kumasi.’
 c. ... [ne-twerefoɔ (*nɔ́)] fi Kumase.
 ... POSS:3SG-writer ART:ANA come.from Kumasi
 ‘... Its author is from Kumasi.’
 Akan (primary data)

- (273) seesia na wɔ̀-fri Nkran be-du-ue. [baase *(nɔ́)] ye-ε lati paa.
 now FOC 3PL-leave Accra FUT-reach-PST bus ART:ANA be-PST late very
 ‘They just arrived from Accra. The bus was very late.’
 Akan (primary data)

The contexts that *nó* was shown in so far are all expected contexts for an anaphoric article. Interestingly, Bombi (2018) and Bombi et al. (2019) show that *nó* in Akan can in fact mark situationally

unique referents. In (274), the referent of *àtààdéé* ‘dress’ is situationally unique; it is similar to the contexts shown in (265) and (266), where *no* was shown to be able to mark a referent as situationally unique together with a pointing gesture, establishing deictic reference. In (274), no pointing gesture is required for *nó* to express situational uniqueness.¹⁸

- (274) Mè-pè [àtààdéé **nó**].
 1SG-like dress **ART:ANA**
 [Context: Kofi and Amma are in the market. Amma disappears and comes back with one dress in her hands. Kofi says:]
 ‘I like the dress.’
 Akan (Bombi 2018: 150)

The other example given in Bombi et al. (2019) to show that the article *nó* is used to mark situationally unique referents is (275). The authors mention that the referent of *àkwàdàá* ‘child’ is unambiguously identifiable solely based on the immediate discourse situation without having been previously mentioned. In fact, this presumably also holds for the referent of *m̀pànyìnfóó* ‘elders’, which is marked by the article *nó* as well. In the case of *àkwàdàá* ‘child’, Bombi et al. (2019) note that the use of the article is required. Because *nó* can and must be used to express situational uniqueness, Bombi et al. (2019) analyse the marker as a definite article.

- (275) [Àkwàdàá #(**nó**)] r̀e-sú níí [m̀pànyìnfóó **nó**] bé-twéń ámà nó ágyàè.
 child **ART:ANA** PROG-cry so elders **ART:ANA** FUT-wait give 3SG stop
 [Context: Badu goes to a naming ceremony [...]. After the libations, everybody is waiting for the next step to happen. A relative of the child approaches Bediako. They haven’t spoken before or talked about anyone. The relative says:]
 ‘The child is crying, so the elders will wait until he calms down.’
 Akan (Bombi et al. 2019: 190)

However, Akan *nó* does not occur with contextually unique referents. The latter are unambiguously identifiable because they can be constructed as unique within a larger context, the latter being implicit. Example (276) shows such a context, where the anaphoric article *nó* cannot be used. The referent of *gúá* ‘market’ is unambiguously identifiable within the larger context of the village or town that the discourse situation is placed in. Importantly, this larger context neither corresponds to the immediate discourse situation nor was it previously referred to. As can be seen in (276), the article *nó* is not used to mark the referent as contextually unique. Another typical example of a contextually unique referent is given in (277); here, we see that the referent of *ɔ́mampanyin* ‘president’ cannot be marked by *nó* either.

¹⁸At this point, it is unclear to what extent this use of *nó* is systematic and subject to inter-speaker variation. It is not unlikely that different varieties of Akan have slightly different conventions for the use of *nó*, resulting in the variation across different sources shown here.

(276) Mò-ró-kò gúá mù.
 1SG-PROG-go market in
 ‘I am going to the market.’
 Akan (Arkoh & Matthewson 2013: 11)

(277) [ɔmampanyin (*no)] hye ntoma-ataadeɛ.
 president ART:ANA wear cloth-dress
 ‘The president is wearing a cloth-made attire.’
 Akan (primary data)

As was mentioned before, the referent of *ɔmampanyin* ‘president’ in (277) is contextually unique because it is a functional concept in the sense of Löbner (1985). It has an inherent possessor argument which evokes another referent, like *president (of Ghana)*. In addition, this relation between ‘president’ and ‘country’ is a one-to-one relation, meaning that by default, there is only a single president of a given country. Bombi et al. (2019) give the example shown in (278) to provide additional evidence that *nó* is a definite article.

(278) [ɔmànpànyíń nó] dùrù-ù hó ànàdwó ò-ńón dú.
 president ART:ANA arrive-PST there night PL-bell ten
 ‘[Recently, there was a government meeting.] The president arrived at 10pm.’
 Akan (Bombi et al. 2019: 184)

In light of the other examples with contextually unique referents, example (278) should rather be viewed as a rel-bridging contexts (in fact, Bombi et al. (2019) introduce it as a bridging context), where reference to the government is made in the immediately preceding utterance. This is confirmed by the fact that *nó* cannot be used with *ɔmànpànyíń* ‘president’ in the sentence in (278) if uttered out of the blue according to my consultant. Therefore, I argue that *ɔmànpànyíń* ‘president’ can be marked by the anaphoric article *nó* in this context only because it establishes an anaphoric relation to the referent of ‘government’ in the preceding utterance.

The matter is somewhat more complicated, though, showing that rel-bridging may still rely on the contextual uniqueness of the referent in the context given by antecedent. Consider example (279), also provided in Bombi et al. (2019).

(279) [ɔsòáfóó #nó] dùrù-ù hó ànàdwó ò-ńón dú.
 minister ART:ANA arrive-PST there night PL-bell ten
 ‘[Recently, there was a government meeting.] The minister arrived at 10pm.’
 Akan (Bombi et al. 2019: 184)

On the one hand, example (279) confirms that the distinction of rel-bridging and u-bridging, with rel-bridging only relying on an anaphoric relation is difficult at least in certain cases. The fact that contextually uniqueness somehow plays a role in rel-bridging cases is also reflected in the English

translation, in which the use of the definite article is somewhat marked as well. A more detailed typology of bridging contexts would go beyond the purposes of the present study and this section. The important patterns regarding the status of Akan *nó* as an anaphoric or definite article can be summarized as follows: if a larger context is made explicit in the preceding discourse segment like in (278), Akan *nó* does seem to be able to mark a referent as contextually unique. Example (279) is crucial because it shows that contextual uniqueness is indeed involved in such uses. Still, examples (276) and (277) showed that the context needs to be given explicitly, if not, Akan *nó* cannot be used, which is why I analyse it as an anaphoric article with extended uses.

Examples (280) and (281) show that the use of *nó* in typical u-bridging contexts is not felicitous either. This is expected given its behavior with contextually unique referents; u-bridging referents are constructed as contextually unique within a context evoked by a previously mentioned referent. However, one could also argue more carefully in saying that the contexts shown in (280) and (281) are associated with possessive markers in such a strong way that there was simply never any functional pressure or need for the article to extend to these contexts.¹⁹ In (280), the referent of *kón* is a unique referent in the context of *ìgùán* ‘sheep’. The same holds for the referent of *nkyensídan* ‘roof’ and *dan* ‘building’ in (281). Examples (280a) and (281a) show that the possessive marker *ní* is used to express the link between the bridging referents. The anaphoric article *nó*, on the other hand, is shown to be infelicitous in such contexts in (280b) and (281b).

- (280) a. Ìgùán *nó* sò árá mà nyímpá ànán nà wó-dzí-ì [*ní kón*].
 sheep ART:ANA big just COMP person four FOC 3PL-eat-PST POSS neck
 ‘The sheep was so big that it was four people that ate its neck.’
- b. # Ìgùán *nó* sò árá mà nyímpá ànán nà wó-dzí-ì [*kón nó*].
 sheep ART:ANA big just COMP person four FOC 3PL-eat-PST neck ART:ANA
 Akan (Arkoh & Matthewson 2013: 14)
- (281) a. Ye-hu-u dan dadaw bi wɔ ekurasí hɔ [*ní nkyensídan*]
 1PL-see-PST building old ART:EXSPEC at village there POSS roof
 ehodwow.
 worn.out
 ‘We saw an old building in the village; its roof was worn out.’
- b. # Ye-hu-u dan dadaw bi wɔ ekurasí hɔ [*nkyensídan nó*]
 1PL-see-PST building old ART:EXSPEC at village there roof ART:ANA
 ehodwow.
 worn.out
 Akan (Arkoh 2011: 80)

¹⁹This observation relates to the use of possessive markers discussed in Section 4.2. The use of articles (definite, anaphoric, or weak definite) in typical u-bridging contexts such as (280) and (281) may reflect an Indo-European bias, where possessive markers seem more restricted than in many other languages of the world. This calls for a closer typological look at the linguistic devices used in such bridging contexts across the world’s languages.

Another context that definite articles may occur with is the absolutely unique referent. While such contexts provide no argument for or against treating *nó* as a definite article, it would be difficult to argue that *nó* is an anaphoric article if it occurred with absolutely unique referents. However, examples (282) and (283) show that this is not the case. In (282), reference is made to the Pope. As a possible explanation for the incompatibility, Arkoh & Matthewson (2013: 19) mention that *Ègyá krónkrón póp* could be similar to a proper name, and that the latter usually do not combine with *nó*.

- (282) Kwámì nyá-à kràtàà fì-ì Ègyá krónkrón póp hó.
 Kwame get-PST letter from-PST father holy pope there
 ‘Kwame got a letter from the holy father Pope.’
 Akan (Arkoh & Matthewson 2013: 11)

The next three contexts feature the referents of ‘moon’ and ‘sun’. Example (283) and (284) show that the anaphoric article does not occur with such expressions. However, in (285) we see that *nó* can occur with absolutely unique referents. Here, the referent is again part of the immediate discourse situation, which may be the reason for the felicitous use of *nó* rather than the absolute uniqueness of the referent *àwìà* ‘sun’ itself.

- (283) Ámstròŋ nyí nyímpá àà ó-dzì-ì kán tú-ù kó-ò òsirán dè.
 Armstrong is person REL SBJ:3SG-eat-PST first fly-PST go-PST moon top
 ‘Armstrong was the first person to fly to the moon.’
 Akan (Arkoh & Matthewson 2013: 11)

- (284) me-n-hu [ɔsram (*no)] anumerε yi.
 1SG-NEG-see moon ART:ANA evening this
 ‘I will not see the moon this evening.’
 Akan (primary data)

- (285) [Àwìà nó] ré-bò ènné.
 sun ART:ANA PROG-hit today
 [Context: Afia is sitting on a bus, when a woman she doesn’t know sits down beside her.
 The woman says:]
 ‘The sun is shining today.’
 Akan (Bombi 2018: 150)

To conclude, we saw that Akan *nó* is used in all contexts associated with anaphoric articles. Moreover, it was shown to occur in situationally unique contexts and bridging contexts in which its use is arguably warranted by the situational or contextual uniqueness. This is clearly no longer a prototypical use of anaphoric articles. However, we also saw that *nó* cannot be used to mark a referent as contextually unique if the context itself is not made explicit verbally (or if it does not

correspond to the immediate discourse situation). The referential functions of *nó* are summarized in Table 5.8.

Table 5.8: The distribution of *no* in Akan

referential functions	use of <i>no</i>
R _{deictic}	(✓) (265), (266)
R _{anaph}	✓ (261), (262), (263)
R _{recog}	✓ (267)
R _{establ}	(✓) (268), (269), (270), (271)
R _{rel-bridge}	(✓) (278), (272), (273)
R _{sit.unique}	✓ (266), (274), (275), (285)
R _{cont.unique}	✗ (276), (277)
R _{u-bridge}	✗ (280), (281)
R _{abs.unique}	✗ (282), (283), (284)

Thus, *nó* differs from definite articles which are systematically used to mark referents as contextually unique. Therefore, I analyse *nó* as an anaphoric article with extended uses; it may be the case that Akan *nó* is about to develop into a definite article. Calling it an emerging definite article, however, would imply that *nó* necessarily develops into a definite article at some point. Because we do not know whether or not *nó* will develop further, I cautiously treat it as an anaphoric article here.

5.2.4 Anaphoric articles with establishing and recognitional referents

As was mentioned in Section 4.1, anaphoric articles are expected to be used with recognitional and establishing referents, which are still pragmatic definite contexts and can be viewed as a common extension of anaphoric articles (or anaphoric markers) in general, which is also suggested indirectly by [Himmelmann \(1997\)](#) who argues that recognitional and establishing contexts are the ones that allow for the development of a demonstrative into a definite article. So far, we saw that the anaphoric articles in Limbum, Komnzo, and Akan can be used to mark recognitional and establishing referents, with the exception of the anaphoric article in Komnzo that is not used in recognitional contexts.

The following paragraphs will provide a few additional examples of anaphoric articles that are used to establish referents as identifiable. While it certainly cannot be a defining property of anaphoric articles, the use with establishing referents does however seem to be a common property of anaphoric articles across languages. An example from the literature is the anaphoric (or strong) article in German (cf. Section 5.3.1). Described and analysed in detail by [Schwarz](#)

(2009), he shows that the strong form of the definite article, which is what I call the anaphoric article here, is the form used in establishing contexts, as in (286).

- (286) Maria ist [#vom / von dem Mann], mit dem sie gestern verabredet
 Maria is by.ART:WEAK / by.ART:ANA man with whom she yesterday date
 war, versetzt worden.
 had stood.up been
 ‘Maria was stood up by the man with whom she had a date yesterday.’
 German (Schwarz 2009: 280)

Another example of an anaphoric article that is also used to mark establishing referents comes from Bunaq (Timor-Alor-Pantar, Indonesia, Timor-Leste). Example (287) shows the anaphoric marker *ba* with an anaphoric referent. We see in (288) that *ba* is also used to establish the referent of *kokoq* ‘water’ as identifiable in the discourse.

- (287) a. Halaqi o bai buleqen o belis t-olo.
 3PL and thing red and white 3INAN-put.in
 ‘They also put in gold and silver things.’
 b. Bai baqa ru-bul gi-e, kalaq gi-e o r-on gi-e.
 thing DEM:DIST.INAN REFL-head 3-POSS neck 3-POSS and REFL-hand 3-poss
 ‘Those are things for the head, for the neck and for the hands.’
 c. [Bai buleqen o belis ba] tumel minak.
 thing red and white ART:ANA.INAN precious.metal complete
 ‘The gold and silver things are completely of precious metal.’
 (Schapper 2009: 276)
- (288) [Halaqi gi-e il kokoq no ba] Suri Guloq a gue sura.
 3PL 3-POSS water bucket OBL ART:ANA.INAN Suri guloq drink PROSP ask
 ‘Suri Guloq asked to drink some of the water of theirs that was in the bucket.’
 Bunaq (Schapper 2009: 279)

The following examples from Teiwa (Timor-Alor-Pantar, Indonesia) show another case of an anaphoric article used with an establishing referent. The anaphoric article *waal* can be seen marking an anaphoric referent in (289); in (290), it is used to establish a referent in the discourse at its first mention.

- (289) a. ana’ maan si ki uwaad nuk yaa, bif ga’an tu’uk.
 long.time NEG SIM eagle big ART:PRES descend child 3SG knock
 ‘Not long [after that] a big eagle comes down and picks the child.’
 b. [Bif waal] ta ba’an yaa, ...
 child ART:ANA TOP fall-REAL descend ...
 ‘That child falls down, ...’
 Teiwa (Klamer 2010: 431)

(290) ... [dei waal la] na gaal-an pin ma ...
 ... bird ART:ANA FOC 1SG shoot-REAL hold come ...
 ‘... that bird I just shot I bring ...’

Teiwa (Klamer 2010: 410)

We find another anaphoric article in Aguaruna (Chicham, Peru), which exhibits the functional extension from anaphoric to establishing referents as well. Example (291) shows the anaphoric article *nu* with an anaphoric referent, and (292) illustrates that it can be used to establish a discourse referent. Note that with restrictive relative clauses, the article *nu* is cliticized to the relative clause.

(291) a. nuni-kā ihīuqā dii-a-ma nunu paŋkĩ
 do.that-Q:SEQ.3:SS discover.PFV:SEQ.3:SS look-IPFV-SUB.NMLZ ART:ANA boa
 hu-ĩ akapi-numa utu-kā akapi-na yu-hu-a-kũ ...
 PROX-LOC liver-LOC enter-Q:SEQ.3:SS liver-ACC eat-APPL-IPFV-SIM.3:SS ...
 ‘Having done that, having discovered (the boa), as they were looking at it, that boa
 having entered here into the liver, it was eating (the man’s liver).’

b. dii-a-ma dukapi asã [nu-na akapi-na] yu-hu-a-kũ
 look-IPFV-NMLZ enough COP:SUB.3:SS ART:ANA-ACC liver-ACC eat-APPL-IPFV-SIM.3:SS
 utu-kā puhu-tai.
 go.in-Q:SEQ.3:SS live-SUB:DS
 ‘As they were looking at it, after enough time it had gone in to eat the liver.’

Aguaruna (Overall 2007: 557-558)

(292) [wi aintsu-na waina-ka-ma-ha=nu-ka] tʃapi-numa puhu-u-ai.
 1SG person-COP:3 see-Q-RECPST-1SG=ART:ANA-FOC Chapi-LOC live-REL-COP:3:DECL
 ‘The person that I saw lives in Chapi [village].’

Aguaruna (Overall 2007: 259)

Evidence for the extension of anaphoric articles to recognitional uses is not as ample as for its extension to establishing ones. However, this may be due to a bias in language descriptions. There are a few examples of anaphoric articles used as recognitional markers, which suggests that this use of anaphoric articles or markers in general may not be uncommon. For instance, Arrernte (Pama-Nyungan, Australia) has the marker *nhenge*, which is glossed as ‘remember’ in the grammar (Wilkins 1989: 121), and which is also mentioned by Himmelmann (1997: 69) as an example of a recognitional marker. Because it occurs in both anaphoric and recognitional contexts, I would treat it as an anaphoric marker whose use is extended to recognitional contexts. Wilkins (1989: 121) describes the use of *nhenge* as follows:

The demonstrative *nhenge* ‘remember’ indicates that the entity to which the NP refers is *something from before which I (the speaker) think that you (the addressee) should be able to remember*.

It often functions as an indicator that something has been mentioned previously in the discourse, although it might not have been mentioned recently. However, it can also be used when something is to be remembered from general context even if there has been no previous mention of it. (Wilkins 1989: 121)

Wilkins provides the examples shown in (293) and (294) to illustrate the two uses. In (293), *nhenge* marks an anaphoric referent (not adnominally in this example), while example (294) demonstrates that *nhenge* can also encode a recognitional referent.

(293) ... kem-irre-ke thiþe kngerreþenhe anteme. kem-irre-me-le ante
 ... get.up-INCH-PST.COMPL bird big.one now get.up-INCH-NPST.PROG-SS and
nhenge alkere-k-irre-ke.
 ART:ANA.NOM sky-DAT-INCH-PST.COMPL
 ‘... A big bird arose. It arose and the aforementioned (bird) took flight.’
 Arrernte (Wilkins 1989: 121)

(294) [inspector **nhenge**] mape-le school nhenhe-rlke inspect-em.ile-ke.
 inspector ART:ANA PL-ERG school DEM-too inspect-TR.CAUS-PST.COMPL
 ‘Those inspectors (you remember the ones) inspected this school too.’
 Arrernte (Wilkins 1989: 121)

Another anaphoric article from the sample that is attested in recognitional contexts is the one from Mongsen Ao (Ao Naga, India). Coupe (2007: 111) explicitly mentions this use (without calling it recognitional), providing the example shown in (295).

“While its most common function is to mark the noun phrases of recurring referents in narrative texts, the anaphoric nominal demonstrative can also be used to determine noun phrases in ordinary everyday conversation if the speaker believes that the interlocutor(s) can uniquely identify a referent from a remembered event or shared knowledge.” (Coupe 2007: 111)

(295) [hmapa təmáŋ sə] tʃhəniŋ-aŋ.
 work all ART:ANA do.work-IMP
 ‘Do all that work [i.e. that we have previously discussed].’
 Mongsen Ao (Coupe 2007: 111)

To sum up, anaphoric articles are systematically and mainly used to mark anaphoric contexts, but their distribution suggests that they are often also used to establish identifiable referents in the discourse and to mark them as identifiable based on shared knowledge outside of the current discourse situation. To what extent these latter two functions are generally available will have to be examined in more detail in future work.

5.3 Weak definite articles

As was mentioned in Section 3.4, Schwarz (2009) introduced an important distinction between strong and weak definite articles, which was taken up in various other studies (Arkoh & Matthewson 2013, Ingason 2016, Irani 2019, Jenks 2018, Ortman 2014, Schwarz 2013, 2019, Šereikaitė 2019).²⁰ Strong definite articles correspond to what I discussed as anaphoric articles in the previous section. Weak definite articles are articles that only occur with semantic definite referents, i.e. they only mark referents as identifiable based on their uniqueness. Dryer (2014) distinguishes between what he calls anaphoric and nonanaphoric definite articles as well, with nonanaphoric articles corresponding to weak definite articles. In this case, I follow Schwarz (2009) in calling such articles weak definite articles. Labeling them non-anaphoric articles is misleading in that it suggests that such markers could also occur in e.g. deictic contexts.

In this section, I will present and discuss the weak definite articles in languages mentioned in the literature, beginning with the languages that provide convincing evidence for weak definite articles in Section 5.3.1. These are the weak definite articles in Fering and German, described in detail by Ebert (1971a,b) and by Schwarz (2009). Sections 5.3.2 to 5.3.5 will then briefly examine the evidence for weak definite articles in Hausa, Lakota, Urama, and Mad'i. I will argue that the data provides rather little evidence for weak definite articles in the strict sense.

5.3.1 Fering and German

Fering is a North Frisian language, spoken on Föhr in Germany by approximately 1500 speakers (Bohn 2004). Ebert (1971a) discusses two articles in Fering, which correspond to anaphoric and weak definite articles. She calls them “A-article” and “D-article” on the basis of their exponents, shown in Table 5.9 (Ebert 1971b: 159):

Table 5.9: Definite articles in Fering

	MASC.SG	FEM.SG	NEUT.SG	PL
A-article	a	at	at	a
D-article	di	det	det	dön

Examples (296) and (297) serve as an illustration for the difference between the two articles *a* and *di* in Fering. In (296), the referent of *sarkkooken* ‘church bells’ is unique in a larger context with a

²⁰Already Lyons (1999: 53-54) notes that Lakota and Hausa have two different definite articles. In the case of Hausa, he notes that both are anaphoric articles in that both occur mostly in anaphoric contexts. For Lakota, he mentions a broader definite article and a more restricted anaphoric article. However, Lyons (1999: 54) also mentions that the anaphoric article does not entirely block the use of the definite article in anaphoric contexts.

single relevant church. The weak article *a* is used in this case. Example (297) shows an anaphoric referent which is marked as such by the use of the anaphoric article *di*.

(296) [A sarkkooken] ringd jister inj.
 ART:DEF_{WEAK} church.bells rang yesterday night
 ‘The church bells rang last night.’
 Fering (Ebert 1971a: 83)

(297) Oki hee an hingst keeft. [Di / *a hingst] haaltet.
 Oki has a horse bought ART:ANA / ART:DEF_{WEAK} horse limps
 ‘Oki has bought a horse. The horse limps.’
 Fering (Ebert 1971b: 161)

A similar distinction between anaphoric articles and weak definite articles is attested in a many German varieties (besides Standard German); Schwarz (2019) mentions the following ones: different Rhineland varieties (Hartmann 1967, Heinrichs 1954), the Mönchengladbach dialect (Hartmann 1982), the Cologne dialect (Himmelman 1997), Bavarian (Scheutz 1988), Austro-Bavarian (Brugger & Prinzhorn 1996, Wiltschko 2013), Viennese (Schuster & Schikola 1984), and Hessian (Schmitt 2006). Furthermore, Ingason (2016) shows that Icelandic, a North Germanic language, distinguishes between a weak definite and an anaphoric (or strong definite) article in the presence of evaluative adjectives in the noun phrase expressing the referent (cf. Section 9.1.1). Here, I will focus on Fering discussed in Ebert (1971a,b) and present it alongside with the weak definite article in German (Schwarz 2009). I will not discuss the morphological details of the Standard German forms of the weak definite and anaphoric articles here; for more details see Schwarz (2009) and references therein.

To illustrate the phenomenon of two definite articles in German, examples (298) and (299) show a pair of contextually unique and anaphoric referents parallel to the ones shown above for Fering. Again, we see that different types of articles, namely the weak definite article in (298) and the anaphoric article in (299) have to be used.

(298) Der Empfang wurde [vom / #von dem Bürgermeister] eröffnet.
 ART:DEF reception was by.ART:DEF_{WEAK} / by ART:ANA mayor opened
 ‘The reception was opened by the mayor.’
 German (Schwarz 2009: 40)

(299) a. In der New Yorker Bibliothek gibt es ein Buch über Topinambur.
 in ART:DEF New York library EXIST 3SG ART:INDEF book about topinambur
 ‘In the New York public library, there is a book about topinambur.’

- b. Neulich war ich dort und habe [**in dem** / **#im** Buch] nach
 recently was I there and have **in** ART:ANA / **in**.ART:DEF_{WEAK} book for
 einer Antwort auf die Frage gesucht, ob ...
 ART:INDEF answer to ART:DEF question searched if ...
 ‘Recently, I was there and searched the book for an answer to the question of [whether
 one can grill topinambur].’
 German (Schwarz 2009: 30)

Establishing contexts are another type of pragmatic definite contexts. As is shown in (300), they need to be marked by the anaphoric article in German, while the weak definite article cannot be used felicitously with an establishing referent.²¹

- (300) Maria ist [**#vom** / **von dem** Mann], mit dem sie gestern verabredet
 Maria is **by**.ART:WEAK / **by**.ART:ANA man with whom she yesterday date
 war, versetzt worden.
 had stood.up been
 ‘Maria was stood up by the man with whom she had a date yesterday.’
 German (Schwarz 2009: 280)

The same holds for recognitional referents in German. As can be seen in (301), a recognitional referent can be marked by the anaphoric article, while the use of the weak definite article is infelicitous.

- (301) Hast du eigentlich nochmal (‘was) [**#vom** / **von dem** Nachbarn]
 have you actually again something **of**.ART:DEF_{WEAK} / **of** ART:ANA neighbor
 gehört?
 heard
 [Context: Anne and Sarah are neighbors who meet occasionally. A while back, another
 neighbor was arrested, which both Anne and Sarah witnessed. Seeing Sarah at the en-
 trance to the building, Anne asks her:]
 ‘Have you heard anything from that neighbor again?’
 German (primary data)

In (302) and (303), we see that the weak definite article in both Fering and German is required with situationally unique referents. These can also be deictic in the sense that the objects linked to the referents of *pokluad* ‘pencil’ and *Buch* ‘book’ can be visible to both the speaker and the hearer. Importantly, the identifiability is not established via the deictic properties of the referents but via its situational uniqueness. In both examples, the referent in question has to be the only

²¹Schwarz (2009: 68) notes that the weak definite article can be used in constructions with non-restrictive relative clauses. In such contexts, the referent of the head noun, marked by the weak definite article, has to be situationally or contextually unique.

salient referent of its kind in the immediate discourse situation, otherwise, the use of the weak definite article is infelicitous.

- (302) smatst' mi ans [at pokluad] auer?
 throw.2SG me PART ART:DEF_{WEAK} pencil over
 'Can you throw the pencil to me?'
 Fering (Ebert 1971a: 104)

- (303) Das Buch, das du suchst, steht [im / #in dem Glasschrank].
 ART:DEF book REL you look.for stands in.ART:DEF_{WEAK} / in ART:ANA glass.cabinet
 'The book that you are looking for is in the glass-cabinet.'
 German (Schwarz 2009: 39)

However, Ebert (1971a: 103) also shows for Fering that a situationally unique referent can be expressed by the anaphoric article. This is shown in (304). She argues that even though the anaphoric article has a stressed counterpart which should be analysed as a demonstrative, a situationally unique referent occurs with the regular anaphoric article, regardless of whether the utterance is accompanied by a pointing gesture or not (Ebert 1971a: 103). The notation used in (304) signals that the article *dèt* does not take the phrasal stress (which is on *búk*), i.e. we deal with the regular anaphoric article. Ebert (1971a: 104) notes that the demonstrative form, i.e. stressed *dèt* would only be used in a context like (304) if more than one book is present in the immediate discourse situation.

- (304) Deest dü mi ans [dèt búk] auer?
 give you me PART ART:ANA book over
 'Can you hand me the / that book?'
 Fering (Ebert 1971a: 103)

We see a similar situation in German, where deictic referents can also be marked by a stressed form of what appears to be the anaphoric article, as is indicated in example (305) by capital letters (*in DEM*). Crucially, neither the default, unstressed anaphoric article *in dem* nor the weak definite article *im* can be used in this context. This stress difference of the article in German has been noted before, leading other studies to analyse such stressed articles used with deictic referents and situationally non-unique referents as demonstrative.²²

- (305) Hans ist in DEM / #in dem / #im Auto gekommen.
 Hans is in DEM / in ART:ANA / in.ART:DEF_{WEAK} car come.PTCP
 [Context: The speaker is pointing at a car, visible to both the speaker and the hearer.]
 'Hans came in that car.'
 German (adapted from Schwarz 2009: 34)

²²For a more detailed discussion, see e.g. Engel (1988: 535,660); Himmelmann (1997: 50-56); Bisle-Müller (1991: 62-66); Hoffmann (2009: 312).

Returning to the weak definite articles in Fering and German, the previous examples showed that they are used to encode situationally unique referents. Moreover, they are only compatible with deictic referents as long as they are also situationally unique; if not, neither the weak definite article nor the anaphoric article can be used. While this goes beyond the scope of the present study, these contexts provide additional evidence for the need to distinguish between deictic and situationally unique contexts. Additionally, the use of the anaphoric article in Fering with situationally unique referents as in example (302) mirrors the use of the anaphoric article in Akan in such contexts (cf. Section 5.2.3). This provides further evidence for situational uniqueness being in an intermediate position between pragmatic and semantic definite functions.

Another referential function that the weak definite articles in Fering and German are expected to express is contextual uniqueness. This is shown in examples (306) and (307) for the articles of both languages. In addition, we see in (307) that the German anaphoric article cannot be used in such contexts. Example (308) shows the same pattern for absolutely unique referents in German: they are necessarily marked by the weak definite article.

- (306) [A köning] kaam tu bischük.
 ART:DEF_{WEAK} king came to visit
 ‘The king came for a visit.’
 Fering (Ebert 1971a: 83)

- (307) Die Prinzessin kam [zum / #zu dem König].
 ART:DEF princess went to.ART:DEF_{WEAK} / to ART:ANA king
 ‘The princess went to the king.’
 German (primary data)

- (308) Armstrong flog als erster [zum / #zu dem Mond].
 Armstrong flew as first to.ART:DEF_{WEAK} / to ART:ANA moon
 ‘Armstrong was the first one to fly to the moon.’
 German (Schwarz 2009: 40)

Turning to u-bridging contexts, the contextual uniqueness of the referent should call for the use of the weak definite article. This is indeed what we find, as is shown in (309) and (310) for both languages.

- (309) a. Wi foon a sark uun a maden faan’t taarep.
 we found ART:DEF_{WEAK} church in ART:WEAK middle of.the village
 ‘We found the church in the middle of the village.’
 b. [A törem] stän wat skiaf.
 ART:DEF_{WEAK} tower stood a.little crooked
 ‘The tower was a little crooked.’
 Fering (Ebert 1971a: 118)

- (310) Der **Kühlschrank** war so groß, dass der Kürbis problemlos
 ART:DEF **fridge** was so big that ART:DEF pumpkin without.problem
 [**im** / #**in dem** Gemüsefach] untergebracht werden konnte.
 in.ART:DEF_{WEAK} / **in** ART:ANA **crisper** stowed be could
 ‘The fridge was so big that the pumpkin could easily be stowed in the crisper.’
 German (Schwarz 2009: 158)

In rel-bridging context, on the other hand, like producer-product relations, weak definite articles are not used. This is shown for Fering and German for the relations between a painting and its painter in (311), and between a theater play and its author in (312).

- (311) a. Peetji hee uun Hamboreg an **bilj** keeft.
 Peter has in Hamburg ART:INDEF **painting** bought
 ‘Peter bought a painting in Hamburg.’
 b. [**Di** mooler] hee ham an guden pris maaget.
 ART:ANA painter has him ART:INDEF good price made
 ‘The painter made him a good deal.’
 Fering (Schwarz 2009: 62)

- (312) Das **Theaterstück** missfiel dem Kritiker so sehr, dass er in seiner
 ART:DEF **play** displeased ART:DEF critic so much that he in his
 Besprechung kein gutes Haar #**am** / **an dem** Autor ließ.
 review no good hair **on**.ART:DEF_{WEAK} / **on** ART:ANA left
 ‘The play displeased the critic so much that he tore the author to pieces in his review.’
 German (Schwarz 2013: 542)

Other examples of rel-bridging are contexts in which the antecedent is an event that the bridging referent refers back to. This is shown in (313) and (314) for Fering and German. Note that in both examples, like in the previous product–producer examples above, the weak definite article cannot be used, and the anaphoric article has to be used instead.

- (313) a. Förgis juar san ik troch Persien an Afghanistan **raaiset**.
 last year am I through Persia and Afghanistan **traveled**
 ‘I traveled through Persia nad Afghanistan last year.’
 b. Ik wal jam fertel, wat ik üüb [**det** raais] ales bilewet haa.
 I wat you tell what I on ART:ANA trip all experienced have
 ‘I want to tell you what I experienced on the trip.’
 Fering (Ebert 1971a: 108)
- (314) a. Hans ist gestern in die Staaten **geflogen**.
 Hans is yesterday in the states **flown**
 ‘Hans flew to the States yesterday.’

- b. **#Beim** / **Bei dem** Flug ging allerdings einiges schief, so dass
at.ART:DEF_{WEAK} / **at ART:ANA** flight went however several.things askew so that
er mit ziemlicher Verspätung am Zielort ankam.
he with quite delay **at.ART:DEF_{WEAK}** destination arrived
‘However, several things went wrong with the flight, so that he arrived with quite a
bit of a delay at his destination.’
German (Schwarz 2009: 279)

The distribution of the weak definite article in Fering is summarized in Table 5.10. The examples of this section showed that although it is compatible with deictic referents, the weak definite article cannot be used to mark pragmatic definite referents, i.e. deictic, anaphoric, and rel-bridging referents. Given this distribution, one would assume that it does not occur with recognitional or establishing referents, either, but this is unclear at this point. In semantic definite contexts, on the other hand, the examples showed that the weak definite article in Fering is systematically used with situationally and contextually unique as well as u-bridging referents.

Table 5.10: The distribution of the weak definite article in Fering

referential functions	use of <i>a/at</i>
R _{deictic}	(✓) (302)
R _{anaph}	✗ (297)
R _{rel-bridge}	✗ (311), (313)
R _{sit.unique}	✓ (302)
R _{cont.unique}	✓ (296), (306)
R _{u-bridge}	✓ (309)

The distribution of the German weak definite article is shown in Table 5.11. Like the Fering weak definite article, it does not occur with pragmatic definite contexts, albeit being compatible with deictic referents. For German, we also saw that the weak definite article is not used with recognitional and establishing referent, where the anaphoric article is used instead. With semantic definite referents, on the other hand, the use of the German weak definite article was shown to be required.

Table 5.11: The distribution of the weak definite article in German

referential functions	use of <i>vom/beim</i> ...
R _{deictic}	(✓) (303), (305)
R _{anaph}	✗ (299)
R _{recog}	✗ (301)
R _{establ}	✗ (300)
R _{rel-bridge}	✗ (312), (314)
R _{sit.unique}	✓ (303)
R _{cont.unique}	✓ (307)
R _{u-bridge}	✓ (310)
R _{abs.unique}	✓ (308)

5.3.2 Hausa

Hausa, a Chadic language with approximately 30 million speakers, is spoken in the Northern region of Nigeria extending to the South of Niger. Moreover, it is one of the official languages of Nigeria and serves as a lingua franca in the whole region of West Africa.

Schwarz (2013) mentions Hausa with another candidate of a weak definite article besides another anaphoric article. Although being careful in indicating that the exact relation of the two markers or articles is not entirely clear, Hausa is mentioned in subsequent work as a potential example of languages with a weak and strong definite article (e.g. Jenks 2018, Schwarz 2019). Therefore, this section will briefly discuss the data in the light of the broader typological overview of this study. To foreshadow the conclusion, the data suggest that Hausa does not have a weak definite article. On the contrary, what would correspond to a weak definite article in Schwarz’s account rather qualifies as an anaphoric article.

The article in question is realized as a segmental suffix with a floating low tone; i.e. as -`r̄ with feminine nouns ending in -a(a) or as -`n with other feminine, masculine, or plural nouns. For the sake of simplicity, I will refer to this marker as -n in the text. The other, potential strong definite article in Schwarz’s account is the marker *dîn*. This marker has indeed been described as an anaphoric determiner or demonstrative (Buba 1997, Jaggar 2001).

For the marker -n, the potential weak definite article in Schwarz (2013), Newman (2000: 143) uses the term “previous reference marker”, describing its use as follows:

I have adopted the familiar term *definite article* instead of the more cumbersome *previous reference marker*. Semantically, however, the latter designation is probably more accurate because it focuses on the language internal (as opposed to the real-world knowledge) determinant of its use. Thus, when a definite NP occurs for the first time in a text, one normally does not use

the d.a. [definite article], even if it would require a definite article in English. (Newman 2000: 143)

Thus, Newman’s description actually rather fits the description of an anaphoric article. Also Jaggar (2001) notes:

The felicity conditions on the use of the DD are difficult to characterize with any precision (Jaggar 1983: 389ff., 1985: 149ff.), but generally speaking its selection is licensed by the presumed unique identifiability of the constituent, usually a noun, to which it is attached, either because the referent has been previously mentioned in the discourse—hence the traditional label “Previous Reference Marker”—or is context-inferable. (Jaggar 2001: 317)

Two examples of *-n* in anaphoric contexts are shown in (315) and (316). Note that I gloss the marker as an anaphoric article, since its distribution and descriptions suggest that it is systematically used with anaphoric referents but not with contextually unique referents.²³

(315) a. Su-na nan zaune, sai ga cinya-r mutum ta fado
 they-IMPV there.LOC sit.STAT then behold thigh-LNK man.SBJ it.PFV fall
 kasa tim, har da jini.
 to.the.ground.LOC bump even with blood.ASSOC
 ‘They were sitting there when a man’s thigh fell to the ground with a bump, blood and all.’

b. Yusha’u ya dubi cinya-r, ya ce ...
 Yusha’u.SBJ he.PFV look.at thigh-ART:ANA he.PFV say ...
 ‘Yusha’u looked at the thigh and said ...’
 Hausa (Jaggar 1985: 167)

(316) a. Shikenan sai ya yi sa’a, ga wasu yara sun zo daidai
 OK then he.PFV do luck behold INDEF boys.SBJ they.PFV come exactly
 guri-n, su-na wasa.
 place.LOC-ART:ANA they-IPFV play
 ‘OK he was lucky, some kids came right to the place, playing around.’

b. Sai yara-n suka zo suka taimake shi ...
 then boys.SBJ-ART:ANA they.PFV come they.PFV help him ...
 ‘Then the kids came and helped him ...’
 Hausa (Jaggar 1985: 166-167)

²³What complicates the distribution of the marker *-r/-`n* even further is that it is also used as a linker in genitive constructions between two nouns. The linker, however, does not have the floating low tone; it is clearly a diachronically related marker, but the article and the linker are not formally identical (Newman 2000: 302), which is not reflected in all examples in this section, since they are not marked for tone. I will not discuss the linker here.

The descriptions *-n* shown above and examples like (315) and (316) strongly suggest that the marker *-n* cannot be a weak definite article. On the contrary, it rather qualifies as an anaphoric article.

The argument in Schwarz (2013) is based on a minimal pair of contexts in Buba (1997: 47), shown in (317) and (318).²⁴ In (317), the referent of *gàri* ‘town’ was not mentioned previously in the discourse. Buba (1997: 43) notes that only *-n* can be used in this context, but not *dîn*. In (318), where the referent of *gàri* ‘town’ is anaphoric, we see the reversed picture with respect to the use of the two markers.

(317) yàayàa gàrî-**n** / #gàrii **dîn**?
 how town-**ART:ANA** / town **DEM**
 [Context: A arrives in B’s place, and he first asks B:]
 ‘How’s the town?’
 Hausa (Buba 1997: 43)

(318) yàayàa gàrii **dîn** / #gàrî-**n**?
 how town **DEM** / town-**ART:ANA**
 [Context: Speaker ass addressee about his journey]
 ‘How’s the town [that you’ve visited]?’
 Hausa (Buba 1997: 43)

Given the distribution of the two markers in (317) and (318), it is of course tempting to analyse *-n* as a uniqueness-based definite marker or as a weak definite article, and *dîn* as a strong definite or anaphoric article. However, as the preceding examples showed, *-n* also occurs in clearly anaphoric contexts, which excludes its analysis as a weak definite article.

As was shown for anaphoric articles from several languages in Section 5.2.4, anaphoric articles can often also mark recognitional and establishing referents. In Akan, the anaphoric article was shown to be even extended to marking situationally unique referents. This could account for the use of the anaphoric marker *-n* in (317), where the referent of *gàri* ‘town’ can be constructed as recognitional or situationally unique. In fact, Jaggard (2001: 318) provides several examples of *-n* “used to encode discourse-new definite referents which are inferable from the extralinguistic situational context”, which I treat as situationally unique. Two examples are shown in (319) and (320), the latter of which could also be argued to show a recognitional article.²⁵

²⁴Buba (1997) does not gloss the examples; the glosses in (317) and (318) are mine.

²⁵Jaggard (2001) does not provide glosses for (319) and (320); the glosses in those examples are mine.

(319) Tô, kàwō kudî-n.
 OK bring money-ART:ANA
 [Context: Said after a bargain has been sealed.]
 ‘OK, give (bring) me the money.’
 Hausa (Jaggar 2001: 318)

(320) Yàyà uwařgidâ-n?
 how wife-ART:ANA
 ‘How’s the wife?’
 Hausa (Jaggar 2001: 318)

Returning to example (318), the incompatibility between *-n* and the anaphoric referent could be due to additional pragmatic restrictions on highly topical referents. In general, anaphoric articles can be in “competition” with other demonstratives in anaphoric contexts just like the definite article in English, depending on the distance to the antecedent and on other intervening referents.²⁶

More evidence for *-n* being an anaphoric rather than a weak definite article comes from its occurrence with establishing referents, typically expressed by a noun with a following restrictive relative clause. Example (321) shows the anaphoric article in such a context, where it occurs with the head noun of the relative clause. Jaggar (2001: 527) mentions that the anaphoric article can additionally occur on the clause-final element.²⁷

(321) Gà [mōtà-ř dà mukà sàyā jiyà].
 PRESENT car-ART:ANA REL 1PL.FOC-PF buy yesterday
 ‘Here’s the car that we bought yesterday.’
 Hausa (Jaggar 2001: 527)

Another establishing context is shown in (322).²⁸ Here, we see in (322a) that *-n* can also occur in the clause-final position, while (322b) shows that the other relevant marker, *dîn*, can occur in this position as well.

(322) a. Kaa san [mutàane-n dà sukà shigoô-n]?
 2SG know people-ART:ANA REL 3PL enter-ART:ANA
 ‘Do you know the people who have come in?’
 b. Kaa san [mutàane-n dà sukà shigoo dîn]?
 2SG know people-ART:ANA REL 3PL enter DEM
 ‘Do you know the people who have come in?’
 Hausa (Buba 1997: 41)

²⁶In fact, Jaggar (1985) is a thorough empirical study of how such factors influence the realization of referential expressions in Hausa. While comparing various types of expressions including nouns with the marker *-n*, he does not include *dîn*, so that a direct comparison of the two markers based on his results is not possible.

²⁷The clause-final use of the article is similar to the clausal determiners seen in Limbum and Akan (cf. Sections 5.2.1 and 5.2.3).

²⁸The gloss in (322) is mine.

As is expected of an anaphoric article, Buba (1997: 42) notes that *-n* cannot be used with a deictic referent with a pointing gesture. Furthermore, he notes that the marker *din* is able to occur with deictic referents if an additional pointing gesture is used.

Another potential complication mentioned by Schwarz (2013: 549) is the fact that previously mentioned referents can be re-introduced without any marker. This observation is taken from Jaggar (2001: 317), who mentions two contexts in which an anaphoric referent is expressed by a bare noun. The two examples shown in (323) and (324) originally come from Jaggar (1985: 155). There, he mentions that 60 separate clauses lie between the referent of *taya* ‘tire’ in (323) and its antecedent. For the referent of *wanzami* ‘barber’ in (324), Jaggar (2001: 155) counts 12 separate intervening clauses including a “potentially ambiguous argument”, meaning that the gap includes another intervening referent. It is difficult to analyse these examples without the larger context, but the long gaps between the antecedent and the anaphoric referent, especially in (323), could account for the absence of the anaphoric article in these cases.

(323) Da muka jawo ta, muka canja taya ...
 when we.PFV pull it we.PFV change tire ...
 ‘When we pulled it [the car] out, we changed the tire ...’
 Hausa (Jaggar 1985: 155)

(324) Da gani-n-sa sai Isa ya tuna da mafarki-n jiya, ya ce
 with seeing-LNK-PART then Isa he.PFV remember with dream-LNK yesterday he.PFV say
 wa wanzami ...
 to barber ...
 ‘On seeing him [the Tuareg], Isa remembered yesterday’s dream and said to the barber ...’
 Hausa (Jaggar 1985: 155)

Jaggar (1985: 168) also mentions that “culturally-prominent referents”, which roughly correspond to what I treat as absolutely unique referents, are expressed by bare nouns in anaphoric contexts. This is shown for *sarki* ‘emir’ in (325).

(325) a. Ran nan su-na zaune da sarki,
 day DEM they-IPFV sit.STAT with emir
 ‘One day they were sitting with the emir,’
 b. sai wani maroki ya zo ya-na ta bunkasa sarki da kirari,
 then INDEF praise.singer he.PFV come he-IPFV keep.on flatter emir with epithet
 ‘when a praise singer came and was flattering the emir with epithets,’
 c. har sarki ya shiga ciki-n abi-n da ya-ke fadi ...
 until emir he.PFV enter inside-LNK thing-ART:ANA REL he-IPFV say.VN ...
 ‘so much so that the emir took notice of what he was saying ...’
 Hausa (Jaggar 1985: 169)

To sum up, the data presented in this section showed that *-n* is not a weak definite article but should better be analysed as an anaphoric article. The other marker, *dîn*, is a demonstrative marker that is also used in certain anaphoric contexts. Without going into the details of the distribution of *dîn*, the important point of the Hausa data is that the presence of an additional anaphoric marker does not necessarily restrict a definite, or in this case, an anaphoric article, to be used in semantic definite contexts that would make it a weak definite article. It may be possible, though, that *dîn* is used more and more frequently over the course of time so that its use in anaphoric contexts becomes systematic, replacing *-n* as an anaphoric article. The latter could also develop further into a definite article, and the fact that we already find it in situationally unique contexts suggest that this is a possibility. Such a development is furthermore facilitated by language contact with English; both Newman (2000) and Jaggar (2001) note that younger speakers and especially Hausa speakers who also speak English are influenced by the English definite article in their use of the anaphoric article *-n* in Hausa. Thus, only in the hypothetical situation that *dîn* develops into a new anaphoric article and *-n* into a definite article, do we have a constellation in which *-n* could be restricted to semantic definites, becoming a weak definite article. Given the data available, however, such a development appears to be entirely hypothetical at this point.

5.3.3 Lakota

Lakota is a Siouan language spoken in the United States and Canada. According to the Endangered Languages Project, Lakota has approximately 2000 speakers.²⁹ Lakota is another language that Schwarz (2013) mentions as a potential candidate for a system with a weak and a strong definite article. The two markers in question are *ki(ŋ)*, the potential weak definite article, and *k'uŋ*, the potential anaphoric (strong definite) article. Similarly to the situation in Hausa, I will show in this section that *kiŋ* should be analysed as a definite article and *k'uŋ* as an additional anaphoric marker.

The relevant contexts that Schwarz (2013) bases his analysis on are given in (326) to (331).³⁰ Examples (326) to (328) illustrate the use of *kiŋ*. In (326), it is used with the absolutely unique referents of *maŋpíya naŋ makhá* 'sky and earth'. In (327), Schwarz (2013: 547) mentions that *kiŋ* marks the referent of a "unique axe present in the narrative discourse situation". Note that in both (326) and (327), the definite referents are followed by the demonstrative *lená/lé* (proximal), which makes it difficult to separate the semantic contribution of *kiŋ* from the one of the demonstrative. The third example of *kiŋ*, example (328), shows that it can be used to mark a bridging referent,

²⁹Cf. <http://endangeredlanguages.com/lang/2052>.

³⁰The examples from Buechel (1939) have no glosses; I added the glosses using the English-Lakota dictionary and Lakota grammar by Ingham (2001, 2003).

which Schwarz (2013: 547) classifies as a part-whole context, i.e. u-bridging.³¹ These contexts suggests that *kiŋ* can indeed be used to mark semantic definites, although without context, (326) and (327) are no clear examples of semantic definite expressions, especially in the presence of the demonstrative in addition to *kiŋ*.

(326) *eya' maŋpiya naŋ makhá kiŋ lená tšoká-káğa-pi k'uŋ hé eháŋ*
 well sky and earth ART:DEF DEM.PL first-create-PASS ANA DEM then
ma-tšúŋ-pe ló.
 1SG-born-PASS ASSERT
 'Well, I was born at the time when the sky and earth were created.'
 Lakota (Schwarz 2013: 546 from O'Gorman 2011: 4)

(327) *tókša nazúŋspe kiŋ lé uŋ wa-káŋuŋuŋiŋ kte, eyá kéye'.*
 eventually axe ART:DEF DEM:PROX using 1SG-smash.up IRR said QUOT
 'In due time I will smash it up with the ax, she said.'
 Lakota (Schwarz 2013: 547 from O'Gorman 2011: 8)

(328) *yaŋkháŋ blé waŋ šmá glakíŋyaŋ ħpáya óhuta kiŋ él inážiŋ.*
 then lake ART:EXSPEC deep across.her.way lay shore ART:DEF to stand
 'Then a deep lake lay across her path. He stood at the shore.'
 Lakota (Schwarz 2013: 547 from O'Gorman 2011: 9-10)

The examples illustrating the use of *k'uŋ* in Schwarz (2013) are shown in (329) to (331). Example (329) is suggestive given the translation as “mentioned before”.³² However, all three examples include the demonstrative *he/héna* (neutral distance) besides *k'uŋ*. Again, this makes it very difficult to determine the semantic contribution of *k'uŋ* as an anaphoric article.

(329) *akícita k'uŋ héna*
 soldier ANA DEM.PL
 'the soldiers mentioned before'
 Lakota (Schwarz 2013: 547 from Buechel 1939: 97)

(330) [*Wišimnaye k'uŋ he*] *kaíyuzeya nájiŋ.*
 tax.collector ANA DEM.SG remote stand
 'The (that) publican stood afar off.'
 Lakota (Schwarz 2013: 547 from Buechel 1939: 327)

³¹I would rather treat example (328) as a rel-bridging context, but as was mentioned in the previous sections, the two types of bridging are often not fully distinguishable.

³²Examples (329) and (330) do not have glosses in the source; the glosses given here are mine. In the gloss of example (329), I translate *wišimnaye* as 'tax collector' according to Ingham (2001).

(331) *igľúštarŋ naŋ heháŋl wikhóškalaka kiŋ [wóyapte č’uŋ hená] wičhák’u*
 finish.a.meal and then young.woman ART:DEF leftovers ANA DEM.PL collect.gave
ške’e.

QUOT

‘He finished his meal and then gave the leftovers to the young woman.’

Lakota (Schwarz 2013: 548 from O’Gorman 2011: 10)

Let us return to the question of whether or not *kiŋ* is a potential weak definite article. The use of *kiŋ* as a definite article in Lakota is well attested (Curl 1999, Rood & Taylor 1996, Van Valin 1977, Williamson 1984). Moreover, Curl (1999) and Ingham (2003) explicitly mention that *kiŋ* can also be used anaphorically. This is shown in example (332), which is the beginning of a story. In (332a), the referent of *wikhoškalaka* ‘young woman’ is introduced. She is referred back to six utterances later by using the article *kiŋ* (spelled *kŋ*), as can be seen in (332b).

(332) a. *el wikhoškalaka wŋ wiyŋ wašte čha lila waštekilapi.*
 there young.woman ART:EXSPEC woman good such very loved.her
 ‘There lived a girl who was very beautiful and greatly loved.’

b. *k’eyaš [wikhoškalaka kŋ] echŋ chŋšni yŋkhŋ hechena*
 but young.woman ART:DEF to.do.it she.was.not.willing and.then at.once
wichaša kŋ lila chŋze nŋ el iyotaka
 man ART:DEF very angry and there sit
 ‘She didn’t care to do that, and immediately he became very angry, saying “Do it”.’

Lakota (Curl 1999: 4-5)

Another example of an anaphoric referent can be seen in (333). In (333c), *kiŋ* (spelled as *ki*) is used with a noun to refer back to *witka* ‘eggs’ first introduced in (333) as a nonspecific referent and established as a specific referent in (333b). Example (333b) also contains a contextually unique referent marked by *kiŋ*. The referent of *huŋku* ‘mother’ is unique within the context of a single nest with eggs.

(333) a. *Tok’á wetu ki lejanl hoksila ki wanhiŋkpe ikikcupi na*
 first be.spring ART:DEF then boy ART:DEF arrow take.POSS.PL and
heyatakiya witka ole ai ...
 mountain.LOC egg seek go.3PL ...

‘At the beginning of spring the boys would take their bows and go away from camp hunting eggs ...’

b. *... s’a na witka oc’aje ki oyas’iŋ kinil mnayanpi na tuktektel*
 ... ITER and egg kind ART:DEF all almost collect.PL and sometimes
[huŋku ki] akignag yuŋka k’eš hoŋpi ogeya ac’ab iheyapi.
 mother.his ART:DEF place.POSS lie although nest whole stabbing shoot.PL
 ‘... and would collect almost all types of eggs and sometimes their arrows would pierce the mother bird also, who was hatching the eggs, the arrow going through the nest.’

- c. na [witka ki] nakuŋ tuktektel oyaŋ'ij kaɦuɦugapi iŋ c'ijpi c'aŋna
 and egg ART:DEF also sometimes all break.PL FOC want.PL when
 oyaŋ'ij aglipi na iyohpewic'akiciyapi c'aŋna taŋtaŋtuŋyan glutapi s'a na
 all bring.PL and cook.3PL.BEN when excessively eat.POSS.PL ITER and
 nakuŋ ziŋtkala ol'ota wic'aopl.
 also bird many shoot.3PL
 'And sometimes also, if they wanted to, they brought home **the eggs** all broken and
 cooked them and ate them to excess and shot many kinds of birds.'
 (Ingham 2003: 95-96)

Thus, examples (332) and (333) showed that *kiŋ* occurs with anaphoric referents, we saw *kiŋ* with a situationally unique referent in (327), and examples (333) and (328) showed that it also marks referents as contextually unique. This clearly supports the traditional analysis of *kiŋ* as a definite article.

Turning to *k'uŋ*, Curl (1999) argues that *k'uŋ*, when used, usually marks anaphoric referents that are re-activated, making it a marker of topic-shift rather than a default anaphoric marker or article.³³ Examining the number of utterances between anaphoric referents and their previous mentions for the main four characters of the story “Double-Face tricks the girl”, she finds that *k'uŋ* usually has 10-20 utterances between the two mentions. In comparison, an anaphoric referent marked by *kiŋ* usually has its antecedent within the previous 1-6 utterances (Curl 1999: 6). Based on this difference in the distance between the anaphoric referent and its antecedent for *kiŋ* and *k'uŋ*, she argues that the latter marker is used to re-introduce referents in the discourse and to signal a topic shift.

Example (334) illustrates the use of *k'uŋ* with an anaphoric referent which becomes the new topic of the discourse segment. At the beginning of the utterance, the Double-Face is at the center of attention, it switches briefly to the bridge and then to the beaver before shifting back to the Double-Face. Both shifts are signaled by the use of *k'uŋ*. As in the examples of *k'uŋ* shown above, the marker occurs together with a demonstrative.

- (334) a. nŋ iŋ eya wana [cheyakthŋpi kʔŋ he] aliwachŋ
 and he too now bridge ANA DEM to.step.on.he.tried
 '[The Double-Face stopped at the shore,] and then he too tried to walk on the bridge,'
 b. k'eyaŋ lila ocik'ayela kŋ ŋ iwŋyak ħʔŋhiya u
 but very narrow ART:DEF on.account.of carefully slowly he.came
 'but because it was very narrow he had to walk very slowly, picking his way with care;'

³³Ingham (2003: 93) follows the analysis of Curl (1999) and notes that *k'uŋ* marks “a switch back to a previous topic”.

- c. ch^hke khoh^h [chapala k^h he] thahenat^h g.lužužu ay^h
 so meantime little.beaver ANA DEM from.this.side taking.apart.his.own he.went
 na chokaya u ĥceh^h okapsakya ch^hke m.ni t'a.
 and midway he.was.coming just.then he.caused.it.to.break.in so water he.died
 ‘so the beaver meantime started to undo the bridge at his end. When the Double-Face
 was halfway across, it broke down, dropping the man into the midst of the lake where
 he drowned.’
 Lakota (Curl 1999: 11)

Example (335) shows the sentence following (334), in which the the referent of *wikhoškalaka* ‘young woman’ is marker by *k’uŋ*, as its last mention is 9 clauses away (Curl 1999: 12). In this case, *k’uŋ* does not occur with an additional demonstrative. The referent of *chapala* ‘beaver’, which was briefly in the center of attention in the previous sentence, is marked by the definite article *kiŋ*.

- (335) hechena [wikhoškalaka k^h] [chapala k^h] alokiksoh^h n^h
 immediately young.woman ANA beaver ART:DEF she.took.her.own.up.in.her.arms and
 lila ŋy^h eceku n^h thiweg.na g.licu ...
 very running that.way.always.she.came and into.camp she.came.back ...
 ‘At once the young girl took her pet in her arms and ran hard all the way homeland and
 soon she entered the camp, to safety ...’
 Lakota (Curl 1999: 12)

To sum up, we saw that Lakota *kiŋ* is a definite article, able to mark pragmatic as well as semantic definites. The other marker, *k’uŋ*, appears to be an anaphoric marker that is especially used with referents whose antecedent is further away and to signal a shift of the center of attention. This situation is not very different from the use of the definite article and the distal demonstrative in English: Both are mutually exclusive as well, and the “stronger” demonstrative is used instead of the definite article in anaphoric contexts with a larger distance between the antecedent and the anaphoric referent.

5.3.4 Urama

Another language that comes close to what could be analysed as a weak definite article is Urama. Urama is a Kiwai language from the group of Papuan languages spoken on the Urama island in Papua New Guinea. J. Brown et al. (2016: 1) note that there are around 6000 speakers of Urama, including speakers of neighboring varieties.

Urama does not have articles in the indefinite domain. J. Brown et al. (2016: 23) analyse the marker =*i* as a definite article and note that it is almost always accompanied by the additional anaphoric marker *aro’o* in anaphoric contexts. Thus, one may reanalyse the article system

and treat =i as a weak definite article and the combination of the two markers, *aro'o ...=i*, as an anaphoric article.

Beginning with the distribution of the article =i, examples (336) and (337) show that it marks contextually unique referents. In (336), =i occurs with the contextually unique referent of *moto* 'house', which belongs to a man introduced in the previous sentence. In (337), it marks *nu'a* 'tree' as unique in the given context.

- (336) a. *aro'o go'ot=i tabo p-emidio vadio dubu ata.*
 ART:ANA village=ART:DEF LOC REMPST-live HAB man certain
 'In that village there lived a certain man.'
- b. [Nu *moto=i*] *umu hiro-hia ka.*
 3SG house=ART:DEF dog many-very PRS
 'At **his home** there were a lot of dogs.'
- Urama (J. Brown et al. 2016: 85-86)

- (337) *Kaukua hini tabo iana nu'a=i ka, ata nu'a ata otidioi haka kekai.*
 ok here LOC final tree=ART:DEF PRS other tree other standing NEG side
 'And then, from there that's the last tree; there's no other tree standing beside it.'
- Urama (J. Brown et al. 2016: 90)

In examples (338) and (339), we see that the definite article is also used in bridging contexts. In both cases, the bridging referent was not mentioned in the discourse before. The context in (338) is clearly a u-bridging context with a part-whole relation between the dog and its tail (note also the use of the possessive marker in the English translation), while the context in (339) may rather be an example of a rel-bridging context. This would then suggest that the definite article is in fact a definite and not a weak definite article.

- (338) *Ita umu gema=i ro nu hete vapo=i ha p-ema'ai.*
 then dog big=ART:DEF NOM 3SG dance tail=ART:DEF EMPH REMPST-give
 'And then the big dog gave him a wag of his tail.'
- Urama (J. Brown et al. 2016: 95)

- (339) *Hi'o beha=i ha horo~horo=i ikedui ri.*
 meat only=ART:DEF EMPH bones=ART:DEF PL.OBJ.throw.away COMP
 'Just the meat only – throw away the bones.'
- Urama (J. Brown et al. 2016: 78)

On the other hand, in most cases in the three narratives provided in J. Brown et al. (2016: 76-95), anaphoric referents are indeed marked by both the definite article =i and by what J. Brown et al. (2016: 23) call the anaphoric demonstrative *aro'o*. Because of its systematic use with anaphoric referents and because of its absence with deictic referents, I analyse *aro'o* as an anaphoric article

here. Examples (340), (341), and (342) show anaphoric referents, marked by both *aro'o* and *=i*. Note that (336) above also contains the anaphoric referent *go'oto* 'village', marked as such by the combination of the two articles.

- (340) Nu'a huna ata Iroroma vati kekai ta; [aro'o nu'a=i] modobo ka pe
 tree big some Iroroma place near LOC ART:ANA tree=ART:DEF can PRS canoe
 ededeai ri.
 make COMP
 'There's a big tree near Iroroma's place; **that tree** could make a canoe.'
 Urama (J. Brown et al. 2016: 22)

While in (341b), *umu gema* 'big dog' is marked as anaphoric by the use of both articles, the referent of *umu keke* 'small dogs' is only accompanied by the definite article *=i*. This referent can be analysed as a u-bridging referent, given that they are a part of the referent of *umu to'o* 'all dogs' mentioned in (341a). This would support an analysis of *=i* as a weak definite article, as opposed to *aro'o ...=i* as an anaphoric article.

- (341) a. Ka ata hivio=i nu aro'o umu to'o=i tui ai ta ga'ubo umu ata
 and one day=ART:DEF 3SG those dog lots=ART:DEF among LOC one dog certain
 nu aro'o umu=i nu huna-hia gema ka.
 3SG ART:ANA dog=ART:DEF 3SG big-very big PRS
 'Then one day among all those dogs there was one dog who was a very big dog.'
- b. Ka [aro'o umu gema=i] ro p-ivoroho vadio [umu
 and ART:ANA dog big=ART:DEF NOM REM PST-PL.OBJ.lead HAB [dog
 keke=i] ka nu davarai ivodau=i.
 small=ART:DEF] and 3SG beach PL.OBJ.take=NMLZ
 'And that big dog used to lead the smaller ones and take them to the beach.'
 Urama (J. Brown et al. 2016: 86)

In (342), on the other hand, we see that the definite article *=i* is also used to mark anaphoric referents without the use of the additional marker *aro'o*, which is also why I treat it as a definite article. The referent of *go'ota* 'coconut (tree)' is mentioned once in (342a) (where it is already identifiable due to previous mention). Then, the coconut (tree) is mentioned again twice in (342b); the first mention in (342b) is marked by the definite article only.

- (342) a. Ioro ohu=i tabo kiaukia bomo gema=i ro go'ota=i ahiai ka.
 climb top=ART:DEF LOC enough pig big=ART:DEF NOM coconut=ART:DEF cut PRS
 'He climbed to the top, and then the big pig started cutting the coconut tree down.'

- b. Go'ota=i p-ahiai ita [aro'o go'ota=i] omo'ai ri
 coconut=ART:DEF REMPST-cut then ART:ANA coconut=ART:DEF fall COMP
 a'ai ta, Iroroma imumuai ka go'ota ata=i to.
 NEARFUT LOC Iroroma fly PRS coconut other=ART:DEF ALL
 'He cut that coconut tree down and when it was about to fall, Iroroma flew onto another
 coconut tree.'
 Urama (J. Brown et al. 2016: 89)

The following examples show that the use of the definite article =i to mark a referent as anaphoric is not a single case. Example (343) follows the sentence from (342b) in the story; the referent of *bomo* 'pig', present in both sentences in (342) is mentioned again, and it is marked only by the definite article =i.

- (343) [Bomo gema=i] va ierehe'edioi ta aro'o go'ota=irautu ahiai ka.
 pig big=ART:DEF then turn LOC ART:ANA coconut=ART:DEF with cut PRS
 'Then the pig turned and he started chopping down that coconut, too.'
 Urama (J. Brown et al. 2016: 89)

What exactly the conditions are for an anaphoric referent to be marked by the definite article =i only or by the additional anaphoric article *aro'o* goes beyond the scope of the present study. It seems likely that factors similar to those shown for the variation between *kiŋ* and *k'uŋ* in Lakota may play a role, namely the distance to the antecedent, topicality and agency in the current discourse segment, and topic shifts or topicality at the beginning of a new topic chain.

The important conclusion regarding weak articles is that the marker =i is a definite and not a weak definite article; it clearly is not restricted to the occurrence in semantic definite contexts. In more general terms, the Urama data presented in this section shows that the availability of a systematic anaphoric marker or article does not necessarily block the use of a functionally broader definite article in at least certain anaphoric contexts (cf. Section 9.1.1).

5.3.5 Ma'di

Another language that was related to weak definite articles is Ma'di. Dryer (2014: 238) lists Ma'di as an example of a nonanaphoric definite article. Assuming that this type of article is equivalent to what Schwarz (2009) calls a weak definite article, this section will briefly review the Ma'di data. I will show that in the case of Ma'di, the analysis of both markers in question as articles is critical, also leading to the conclusion that Ma'di does not have a weak definite article.

Ma'di is a Central Sudanic language, spoken in the South of Sudan and in the North of Uganda; Blackings & Fabb (2003: 3) estimate the size of the Ma'di population at approximately 250,000 people. Regarding its article system, Ma'di has what Blackings & Fabb (2003: 17) call an "specific

indefinite determiner”. As will be shown, what is treated as a definite article in Ma’di is likely not a definite article. Only what is described as an anaphoric marker actually qualifies as an anaphoric article.

The markers in question are what Blackings & Fabb (2003) call the definite article *rì* and the anaphoric marker *nā*. They describe the two markers as follows:

“Referents which are to be found in the discourse context are identified with *rì* ‘the ... in question’ (for more distantly mentioned referents) and *nā* ‘the aforementioned’ (for more recently mentioned referents). These two determiners are definite but can become indefinite when they are preceded by a postmodifier.” (Blackings & Fabb 2003: 17).

Throughout the grammar, the so-called definite article often occurs together with the focus marker *ĩ* and also its translation as “the ... in question” and its description given above make it questionable whether the marker is a definite article at all. Example (344) explicitly shows that referents marked by *rì* does not need to be interpreted as identifiable or definite.³⁴

- (344) *ópí ò-dū [gàlámò ìkā rì] ĩ.*
 Opi 3-TAKE pen red DEF FOC
 ‘It is the / a red pen that Opi took.’
 Ma’di (Blackings & Fabb 2003: 17)

Without offering a full reanalysis of *rì*, two properties need to be mentioned that show that it is not a definite article, even though its functions may overlap with those of a definite article. It appears to be similar to clausal determiners in other languages. We thus find *rì* with restrictive relative clauses as in (345) and (346) and it also marks adverbial clauses as topical or backgrounded and given information, as in (347).³⁵

- (345) *[ō-gbì-bá rì] ògū.*
 ITER-shoot-SUB DEF thief.PL
 ‘The ones who (repeatedly) shot at them are thieves.’
 Ma’di (Blackings & Fabb 2003: 22)

- (346) *[àràbià ópí ?à dži-lé rì] pá nā ādī rá.*
 car Opi POSS take-SUB DEF leg AFF deflate AFF
 ‘The car which Opi took has a flat tyre.’
 Ma’di (Blackings & Fabb 2003: 22)

³⁴I nevertheless gloss *rì* as DEF following Blackings & Fabb (2003). Blackings & Fabb (2003: 279) also mention that an indefinite specific marker, a floating low tone, can combine with *rì*. If this happens, they argue that *rì* absorbs the low tone and remains as a sole marker of an indefinite referent. This would explain the alternative indefinite interpretation of *gàlámò* ‘pen’ in (344), but even then one would need to account for *rì* being compatible with an indefinite referent.

³⁵Blackings & Fabb (2003: 279-280) discuss a few alternative analyses of *rì* that point into the same direction.

function of demonstratives (e.g. [Himmelmann 1997](#): 61-72, [Diessel 1999](#): 105-109). As was explained in detail in Section 3.2.3, recognitional referents are mutually and unambiguously identifiable based on shared (personal) knowledge or mutual experience in the in past. Even though [Himmelmann \(1997\)](#) argues that the identifiability of recognitional referents can only be based on personal or specific knowledge as opposed to contextually unique referents which are identifiable on the basis of general or world knowledge, recognitional markers often occur with cultural concepts, i.e. with referents that involve shared common rather than personal knowledge. This will apply to a number of examples presented in this section as well. I will return to this issue in Section 5.4.6.

In the sample, only 5 out of 104 languages feature what I call a recognitional article. The languages are: Lavukaleve (Papuan, Solomon Islands), Oksapmin (Nuclear Trans New Guinea, Papua New Guinea), Bininj Kun-Wok (Gunwinyguan, Australia), Gooniyandi (Bunaban, Australia), and Yankunytjatjara (Pama-Nyungan, Australia). It is obvious that those languages are all spoken in Papunesia and Australia, making it tempting to think of recognitional articles as an areal phenomenon.³⁶ This may indeed be the case, but given that recognitional uses are generally not often mentioned in grammar, it may simply be the result of a different traditions of grammar writing. Therefore, more research is necessary on the expression of recognitional referents and the crosslinguistic availability of dedicated recognitional markers in order to draw any conclusions. The following paragraphs will briefly present each of the 5 recognitional articles in turn.

5.4.1 Oksapmin

Oksapmin is a Nuclear Trans New Guinea language spoken in Papua New Guinea with approximately 8000 speakers ([Loughnane 2009](#): 1). Besides the recognitional article, Oksapmin has a more general definite and an indefinite article.³⁷ According to [Loughnane \(2009: 123\)](#), “[t]he demonstrative *max* is usually used when the referent has not been previously mentioned/activated in the current discourse but is presumed to be familiar to both the speaker and the addressee”. Two examples of the recognitional article are given in (350) and (351). [Loughnane \(2009: 124\)](#) notes for example (350): “In the following example, *max* is used to refer to the story that the speaker told earlier in the morning. This text had not been previously mentioned in the current story but all the addressees had been present when he told the previous story”:

(350) gin i ml-sa jəxe tumbuna paxna sup [stori **max**] pla gina.
 now HESIT come.up-SEQ then ancestor hunger illness story **ART:RECOG** tell.PRS.SG now
 ‘Now, I came up and told that story about famine in the old days. Now ...’

Oksapmin ([Loughnane 2009](#): 124)

³⁶Interestingly, I am not aware of any Austronesian language having a dedicated recognitional marker.

³⁷See Section 9.2 for a discussion of functionally overlapping articles within languages.

The context of example (351) is the following: “The text to which the following example belongs was collected just after New Year’s Day which everybody in the community had known about and the churches had held special events for” (Loughnane 2009: 124):

- (351) [niu jia **max**] bəten x-t-pel=o li=m xe-ja.
 new year **ART:RECOG** pray O-PFV-FUT.PL=QUOT say=SEQ be-PRS.PL
 ‘They wanted to pray for you, you know, that New Year.’
 Oksapmin (Loughnane 2009: 124)

Another example of the use of *max* is given in (352), where it occurs with the first mention of *əxəsan*, the name of a certain variety of birds that the speaker and the hearer are presumably familiar with (Loughnane 2009: 123). Note that in (352), the recognitional article *max* is used to refer to a type of bird, which would probably fall within cultural or general knowledge rather than within shared personal knowledge or experience. Also in (353), *max* occurs with the referent of *ninan*, ‘green leafy vegetable’, which appears to belong to culture-specific knowledge. In fact, quite a few examples of *max* in Loughnane (2009) involve identifiable referents that suggest culture-specific knowledge and not necessarily shared personal knowledge. At this point, it is not clear whether this should count as an extended use of recognitional articles or whether this corresponds to its prototypical use. It also suggests that the two categories of personal or specific knowledge as opposed to general world knowledge are not sufficient or at least not always applicable as such.

- (352) gin mən=a [əxəsan **max**=xe] go den x-pat=d=a.
 now brother=LNK bird.variety **RECOG**=FOC 2SG food O-IPFV.PRS.SG=Q=EMPH
 ‘Now, brother, you know that *əxəsan*, do you like eating (it) as well?’
 Oksapmin (Loughnane 2009: 123)

- (353) xanəp xəp-tu-pa dus jox a [ninan **max**] suxu-m
 person die-PFV-REMPST.PL inside **ART:DEF** HESIT bush.kumu **ART:RECOG** collect-SEQ
 d-sxe.
 eat-REMPST.PL.HAB
 ‘In the midst of the famine, we used to collect and eat that (inedible) *ninan*.’
 Oksapmin (Loughnane 2009: 246)

We do however also find uses of *max* which clearly involve personal knowledge; two examples are given in (354) and (355). The referent of *brokoli* ‘broccoli’ in (355) is not only marked as recognitional but it also established, since the relevant information to unambiguously identify the referent is provided in the clause following the head noun.

(354) nox [natan oxe kol max=a] p-ti-p.
 1SG Nathan POSS:3SG.M sister ART:RECOG=EMPH tell-PFV-REMPST.SG
 ‘I told him, “I’m, you know, Nathan’s sister.”’
 Oksapmin (Loughnane 2009: 248)

(355) toxan a-sxa-pat=xe mǝ=ma gologwe plastik tem
 sweet.potato BEN-look.after-IPFV.PRS.SG=SUB DEM:PROX=REL 2SG.REFL.POSS plastic inside
 mǝ=ma brokoli lin uŋ n-a-sli-m x-t max
 DEM:PROX=REL broccoli leaf a.lot O:1/2-BEN-put-SEQ be-IPFV.HESTPST ART:RECOG
 i=ma lin uŋ jox=si a-p-lu-xu-l.
 DEM:DIST=REL leaf bag DEF=with BEN-CAUS-go.up-PFV-HESTPST
 ‘After I gave him sweet potato, that bag of broccoli leaves which (I saw that) you put in
 your bag for me yesterday, I took that bag of leaves up for him.’
 Oksapmin (Loughnane 2009: 461)

As was shown in Section 4.1, Oksapmin has a separate demonstrative marker to encode deictic referents. Example (356) shows this.

(356) ox mǝ=tǝx pa se gno-m pt-sux=li.
 3SG.M DEM:PROX=place taro INFER grow-SEQ be-HAB.EVID.REMPST=REPORT
 ‘I guess he must have stayed here [this place] growing taro.’
 Oksapmin (Loughnane 2009: 110)

To argue that *max* is a recognitional article and no definite article, examples (357a) and (358) illustrate that *max* is not used with anaphoric and contextually unique referents. In such contexts, Oksapmin employs the definite article *jox*.

(357) a. [ap tit] x-t x-n-gop=li ej
 house ART:INDEF do-PFV-EVID.HODPST.SG be-PFV-VIS.REMPST.SG=REPORT gosh!
 x-ti-p x-n-gop=li je put te put
 do-PFV.EVID.REMPST.SG be-PFV-VIS.REMPST.SG=REPORT mountain top place top
 te=nǝp mox ...
 place=INT DEM:PROX ...
 ‘There was a house which had just been built, sorry, which had been built long ago.
 Right at the very top of that mountain ...’
 b. ...ux [ap jox] loj-xi-p=li
 3SG.F house ART:DEF enter-PFV-EVID.REMPST.SG=REPORT
 ‘She went into the house.’
 Oksapmin (Loughnane 2009: 129)

- (358) tom jox lum p-d-m edi-pla=o [ake jox]
 water ART:DEF a.lot CAUS-eat-SEQ stay.PFV-REMFUT.SG=QUOT stomach ART:DEF
 ox=o tom=wi x-ti-plox=xejox n-pli-nuj.
 2SG.M=QUOT water=only be-PFV.HODFUT.SG=because O:2-tell-EVID.HODPST.SG
 ‘Don’t give her too much water! Her stomach will fill up with water, she told me.’
 Oksapmin (Loughnane 2009: 480)

5.4.2 Lavukaleve

Lavukaleve is another language with a marker that I treat as a recognitional article. Lavukaleve is a Papuan language spoken on the Russel Islands in the Central Province of the Solomon Islands. According to Terrill (2003: 1), the language has approximately 1700 speakers. Besides the recognitional article *mea*, Lavukaleve has a definite and an indefinite article.

While I classify *mea* as a recognitional article, Terrill (2003) argues against such an analysis. She describes the use of *mea* as follows:³⁸

“In constructions with nouns, *mea* marks out an entity as being a particular one, known to both speaker and addressee. It refers to shared knowledge, shared not because it has been referred to already in the discourse, but shared in the sense that everybody knows it. It refers to common knowledge, common sense and shared knowledge of the world, which is based on the fact that all Lavukals live in the same place, and have always done so (at least in their cultural mythology) and share the same culture and world view.” (Terrill 2003: 83-84)

Because of the use of *mea* based on common cultural knowledge rather than specific personal knowledge between a given pair of discourse participants, Terrill (2003: 84-85) explicitly notes that *mea* is not a recognitional marker in the sense of Himmelmann (1997). Furthermore, Terrill (2003: 85) argues that *mea* is not used as a request for the hearer to signal whether or not sufficient information for the identification of the referent was provided by the speaker, which is one of the uses mentioned by Himmelmann (1997) (cf. Section 3.2.3). She notes that “*mea* is instead an acknowledgment of the fact that the speaker and the hearer share their knowledge and experience of the world” (Terrill 2003: 85). As was mentioned in Section 3.2.3 and shown in the previous section for the recognitional article in Oksapmin, such uses are in fact commonly attested with recognitional articles, which is why I treat Lavukaleve *mea* as a recognitional article as well.

Examples (359), (360), and (361) show the use of *mea* to mark a referent as unambiguously identifiable based on shared cultural knowledge. Note that the recognitional article can occur on its own as in (359) and (360), but it is also used together with the definite article as in (361).

³⁸The form *mea* is the feminine singular form of the article; I follow Terrill (2003) in using it as the citation form.

- (359) Ni'-kol ta'vuli-mal fiv ngo-koroi-re-v ta'vuli-mal va [houla lagari
 first house.post-PL 3PL.FOC 2SG-chop-FUT-PL house.post-PL ART:DEF.PL stick choice
 me-a] feo.
 ART:RECOG-SG.F 3SG.F.FOC
 'First you cut the posts. The posts are from this special tree (which we all know).'
 (Terrill 2003: 84)
- (360) La-e-kiu-ge lo-vulima-n [suluverav me-v] ma-laketei
 O:3DU.M-SUB-die-ANT POSS:3DU-behind-LOC chiefs ART:RECOG-PL POSS:3PL-life
 tugua ta oi-i.
 exchange just s:3SG-do
 'When [the two men] died, after them, the lives of these chiefs [who we all know] changed.'
 Lavukaleve (Terrill 2003: 84)
- (361) Mina hori hano [kastom me ga] e-na, ta
 thing.F DEM:PROX.SG.F then custom.N ART:RECOG.SG.N ART:DEF.SG.N O:3SG.N-in time
 a-n siriae-e e-hai ma-me-m na a-na, ...
 O:3SG.M-in catch.fish-NMLZ O:3SG.N-do s:3PL.HAB-SG.M ART:DEF.SG.M O:3SG.M-in ...
 'This thing, in this custom before [which we all know], when they would go fishing, ...'
 Lavukaleve (Terrill 2003: 84)

Terrill (2003: 86) also notes that the recognitional article is commonly used with “proper nouns, in particular, places and people familiar to all Lavukal people”. Examples (362) and (363) show the use of *mea* with two place names.

- (362) [Ma Baesen me-a la] o-soi-re ae lavea
 sea.passage Baesen.F ART:RECOG-SG.F ART:DEF.SG.F O:3SG.F-follow-NONFIN go.up appear
 lo-me ga.
 s:3SG-HAB.SG.N ART:DEF.SG.N
 'It (m.) came along and appeared in the sea passage at Baesen.'
 Lavukaleve (Terrill 2003: 86)
- (363) foa [Nagu me] e-fataran ka
 go.down place.in.Isabel.(N) ART:RECOG.SG.N O:3SG.N-straight.out.from LOC.EMPH
 sia-vel hokari ta fi.
 do-COMPL here.PROX just 3SG.N.FOC
 'They went down, then opposite Nagu they went ashore.'
 Lavukaleve (Terrill 2003: 157)

Note that the use of the recognitional marker is not restricted to proper nouns. In example (364), we see its use with a common nominal expression *kiu* ‘the sick’ to mark the referent as identifiable by the speaker.

(364) [kiu **me-v**] fele-re vo-sou fale-ri.
 die ART:RECOG-PL return-NONFIN O:3PL-rise stand-CAUS
 ‘[we] help the sick to stand up again’ (lit: ‘stand the dying back up’)
 Lavukaleve (Terrill 2003: 238)

Finally, *mea* occurs with referents that are established as identifiable, as is shown in (365). Here, *mem* together with the definite article *na* introduce the referent of *marai* ‘warrior’ and signal that the hearer should construct it as identifiable, even though the information about the warrior’s identity is provided in the following clause.

(365) Aka Lavukale-n hoka [marai **me-m** na]
 then Russels.Is.-LOC here.PROX warrior.M ART:RECOG-SG.M ART:DEF.SG.M
 o-langi fi Todou.
 POSS:3SG-name.N 3SG.N.FOC Todou
 ‘So here in the Russells there was a warrior, his name was Todou.’
 Lavukaleve (Terrill 2003: 192)

For spatial deixis, Lavukaleve uses a different adnominal marker *hoia* (f.) / *heana* (m.) (Terrill 2003: 177-183). Example (366) shows this:

(366) Aka [Suvala **heana**] fin a-ea-re-m hin.
 then Sulava.M DEM:DIST.SG.M 3SG.M.FOC S:1SG-talk.about-FUT-SG.M 3SG.M.FOC
 ‘I’ll talk about that Sulava island over there.’
 Lavukaleve (Terrill 2003: 178)

As was mentioned, Lavukaleve also has a definite article, which marks anaphoric and contextually unique referents (amongst others) as identifiable. Two examples to illustrate these two types of contexts, respectively, are given in (367) and (368). Like the recognitional article and the demonstrative, the definite article inflects for number and gender; it has different exponents with masculine and feminine head nouns.

(367) a. **meo** vo-ha fi ma-me.
 tuna.PL 3PL:O-clear 3SG.N.FOC S:3PL-HAB
 ‘They were catching bonito.’
 b. [meo **va**] vo-e-tegi-ge ma-vau.
 tuna.PL ART:DEF.PL O:3PL-SUB-feed-ANT S:3PL-go.seawards
 ‘When the bonito were feeding, they went seawards.’
 (Terrill 2003: 92)

(368) a. Aka loval sevi fiv ma-langi ga.
 then giant.trevally.PL eight FOC.3PL POSS.3PL-name.N ART:DEF.N
 ‘The eight fishes, they are called loval.’

- b. [Savata-m na] fin fafas hin.
 ninth-SG.M ART:DEF.M FOC:3SG.M fafas.M FOC:3SG.M
 ‘The ninth is a fafas.’
 (Terrill 2003: 77)

5.4.3 Bininj Kun-Wok

Bininj Kun-Wok is a Gunwinyguan language spoken in the Northern Territory of Australia.³⁹ According to Evans (2003: 6), the language has about 2000 fluent speakers, not all of which are native speakers, since the language is used as a lingua franca in the region.

Bininj Kun-Wok does not use articles other than the recognitional article *nawu*, although it has a relatively complex system of demonstratives. The form *nawu* includes the masculine class prefix *na-*; I will use this form in the text to refer to the marker following Evans (2003).⁴⁰ The marker *nawu* is not explicitly treated as a recognitional marker but its use is described as follows:

“Because identification is mediated by labelling rather than pointing or reference, it is never used without some further descriptive material (i.e. independently), which may range from a noun to a relative clause; syntactically, it is therefore always adnominal and never pronominal.” (Evans 2003: 297)

Examples of *nawu* with recognitional referents are given in (369) and (370). The utterance in (369) is the first mention of the high country after a discussion of lowland activities (Evans 2003: 298). We see that the referent of *garrigad* ‘high country’ occurs with the recognitional article *nawu*, marking it as identifiable.

- (369) Well sometimes barri-re-i gonda [garrigad **nawu**] ...
 well sometimes 3.PST-go-PST.IPFV there high.country ART:RECOG.M ...
 ‘When they’d go up into **that high country** ...’
 (Evans 2003: 297-298)

Also in (370), *nawu* is used in the first mention of the referent of *gunj* ‘kangaroo’. Because first referred to by *maih* ‘animal’, the recognitional context in this example suggests that the speaker is unable to come up with the appropriate lexical expression and uses a more general expression first. At the same time, the marker *nawu* codes the referent as identifiable for the speaker based on shared knowledge outside of the immediate discourse situation, requesting the hearer to construct the referent as identifiable nevertheless. This use of *nawu* becomes even more evident in (371),

³⁹Other names are Kunwinjku (Gunwinggu) and Mayali; to be precise, Bininj Kun-Wok includes a high number of different dialectal varieties (Evans 2003: 7).

⁴⁰The form *nawu* is used in the Gundjeihmi variety; other realizations are *nabu*, *ngalbu*, and *ngalu* in the varieties of Kune, Kunwinkju and Manyallaluk Mayali (Evans 2003: 297).

where it is used together with the dedicated element *njamed* ‘whatchamacallit’, which functions as a placeholder in case the speaker cannot come up with the proper expression (cf. Evans 2003: 283).

(370) [Maih **nawu**], [gunj **nawu**] bonj andiwo gunj andud.
 animal ART:RECOG.M kangaroo ART:RECOG.M OK give.IMP kangaroo then
 [Context: Emu has just returned, and these are her first words:]
 ‘Right, that animal, that kangaroo, give it to me then.’
 Bininj Kun-Wok (Evans 2003: 297)

(371) [Njamed **nawu**], ragul nungga gun-dulk ba-me-i.
 whatchamacallit ART:RECOG.M red-eyed.pigeon he CL:IV-stick 3.PST-get-PST.PFV
 ‘That whatchamacallit, the red-eyed pigeon, he picked up a stick.’
 Bininj Kun-Wok (Evans 2003: 298)

Another type of context in which *nawu* is used is shown in (372). The referent of *gunj* ‘kangaroo’ had been a salient discourse participant in the preceding discourse, but not in the immediately preceding discourse segment. Therefore, the speaker might be uncertain whether or not the hearer can successfully establish reference. The speaker utters *gunj* ‘kangaroo’ together with *nawu* as an afterthought, which signals to the hearer that they can unambiguously identify the referent.

(372) Ba-yerrng-yiga-ni ba-djoleng-m-inj ba-ru-y
 3.PST-wood-fetch-PST.IPFV 3.PST-cooked-INCH-PST.PFV 3.PST-cook-PST.PFV
 [**nawu** gunj].
 ART:RECOG.M kangaroo
 ‘While she (emu) was getting wood it got cooked and ready, that kangaroo.’
 Bininj Kun-Wok (Evans 2003: 298)

The *nawu* marker series is also used as a relative clause marker in Bininj Kun-Wok (Evans 2003: 642-646). Two examples are given in (473) and (374) :

(373) a. Galuk nagudji [**nawu** ba-di gu-mekke], [**nawu** gare yirrdjdja
 then one.M ART:RECOG.M 3.PST-stand LOC-there ART:RECOG.M maybe Yirrdjdja
 o duwa, njamed],
 or Duwa what
 ‘Then one [youth] who was right there, who was maybe Yirrdjdja or Duwa, whatever,’
 b. ba-bili-doh-dombu-ni dja ga-warre, barri-djal-ru-ngi.
 3.PST-fire-INCEP-extinguish-PST.IPFV and 3-bad 3AUG.PST-just-burn-PST.IPFV
 would try to put out the fire but not to avail, they’d all just get burned.
 Bininj Kun-Wok (Evans 2003: 643)

the recognitional article *panya*. Although treated as an anaphoric marker in Goddard (1985: 54), the description of its use corresponds to the marking of recognitional reference.

panya [...] (roughly ‘you know the one’) calls the listener’s attention to the fact that he or she is already familiar with a referent. It is not usually used about things which are fully topical– i.e. already being talked about, but rather to re-introduce something into the conversation. (Goddard 1985: 54)

Two examples of *panya* marking a recognitional referent are shown in (378) and (379). In both cases, it is used to mark the referent it occurs with as identifiable by the hearer based on shared knowledge and experience and not because of previous mention. Goddard (1985: 54) notes that in (379) “a group of hunters return to camp to find it unoccupied, but then see smoke on the horizon and recall that there had been talk of checking a nearby waterhole”.

(378) ayi, [pika **panya** ngayulu] wiya-ri-ngu ngaltutjara, [mititjini
hey pain **ART:RECOG** 1SG nothing-INCH-PST dear.thing medicine
panya-ngka-lta ngayulu palya-ri-ngu.
ART:RECOG-LOC-and.then 1SG good-INCH-PST
‘Hey, you know that sickness of mine has disappeared. The dear (stuff), with that medicine,
you know the medicine, I got better.’
Yankunytjatjara (Goddard 1985: 54)

(379) munu nyakupayi “waru kampa-nyi [kapi **panya-ngka-manti-r**]!”
ADDIT see fire.NOM burn-PRS water **ART:RECOG-LOC-probably-excitement**
‘And they would see “There’s a fire burning, at that water(hole), you know the water(hole),
most likely, by gee.” ’
Yankunytjatjara (Goddard 1985: 54)

Example (380) shows a context in which the referent of *waru* ‘firewood’, marked by the recognitional article *panya*, is uttered as an afterthought. Yankunytjatjara has a verb-final word order; Goddard (1985: 21) notes that postverbal nominal expressions can “have ‘afterthought’ status (eg confirming the identity of a referent taken for granted in the preceding stretch of the utterance), [...]”. As was shown in example (372) for the recognitional article in Bininj Kun-Wok, using *panya* with the noun expressing the referent confirms the identity of the referent, at the same time signaling that the hearer can indeed unambiguously identify the referent.

(380) manta-ngka wani-ngu [waru **panya**].
ground-LOC throw.away-PST fire.ACC **ART:RECOG**
‘[He] threw it on the ground, that firewood.’
Yankunytjatjara (Goddard 1985: 41)

Examples (381) and (382) show *panya* in contexts where the referent is first expressed by a more general nominal expression marked by *panya*, which signals that the hearer is able to identify the referent. However, to ensure that the hearer can unambiguously identify the referent, the speaker then apposes another referential expression with more lexical content.

(381) kaa [katja panya], inkilyi-ku katja, wati waputju-ngka
 CONTR SON.NOM ART:RECOG co.parent.in.law-GEN SON.NOM man father.in.law-LOC
 wangka-nytja wiya.
 talk-NMLZ NEG
 ‘But that son, the co-parent-in-law’s son, doesn’t talk to (his) father-in-law.’
 Yankunytjatjara (Goddard 1985: 124)

(382) paluru wangka-ra ara tjakultju-nu, report, [mitingi panya] Pitjanytjara
 DEF.ERG talk-SER matter.ACC inform-PST report.ACC meeting ART:RECOG Pitjanytjara
 Council miting-ka mungalyuru.
 Council meeting-LOC in.future
 ‘He’ll report the matter, to the meeting, the Pitjantjatjara Council meeting, soon afterwards.’
 Yankunytjatjara (Goddard 1985: 131)

In (383), we see that the recognitional article *panya* in Yankunytjatjara can also be used to establish a referent as identifiable by the hearer. Similarly to the use as a restrictive relative clause marker in Bininj Kun-Wok, Goddard (1985: 77-80) describes the use of *panya* as a relative clause marker. Another example of *panya* in this function is given in (384).

(383) nyara-kutu kati-ma, [ngura panya] puyu paka-ntja-la-kutu.
 over.there-ALL take-IMP.IPFV place ART:RECOG smoke.NMLZ come.up-NMLZ-LOC-ALL
 ‘Take it over there, to that place (where) smoke’s coming up.’
 Yankunytjatjara (Goddard 1985: 79)

(384) [wati panya] mungartji ngalya-yanku-nytja-lu mutaka kati-ngu.
 man ART:RECOG yesterday this.way-go-NMLZ-ERG car.ACC take-PST
 ‘The man (who) came yesterday took the car.’
 Yankunytjatjara (Goddard 1985: 78)

Another important point mentioned by Goddard (1985) is that the frequent use of the recognitional article may go back to certain cultural factors, namely that the use of names is generally avoided. He notes:

“Yankunytjatjara people like to avoid referring to familiar individuals by name, which is in any case largely unnecessary in a small kin-based community where people all know each other and their mutual relationships. So one often hears things like *wati panya yaaltji?* ‘where’s that bloke, you know the one?’ ” (Goddard 1985: 54)

The recognitional article *panya* is not used with deictic referents. Goddard (1985: 53) lists the following demonstratives for Yankunytjatjara: *nyanga* ‘this’, *pala* ‘this just here/here’, and *nyara* ‘that over there’. An example of a deictic referent, expressed by *nyanga*, is given in (385).

- (385) [wati **nyanga-ngku**] papa pu-ngu.
 man **this-ERG** dog.ACC hit-PST
 ‘This man hit the dog.’
 Yankunytjatjara (Goddard 1985: 53)

The recognitional article is not used (at least not systematically) with anaphoric referents either. One example is given in (386), where the anaphoric referent of *tjina* ‘track’ does not have to be marked as anaphoric when mentioned again in (386b).

- (386) a. watja-ni-lta “uwa **tjina**-la yanku-la nya-wa”, ...
 say-PRES-and.then yes **track.ACC-1PL.ERG** go-SER see-IMP ...
 ‘He’s saying “Yes, let’s follow the tracks”, ...’
 b. wana-ra-mpa **tjina** wana-ra-mpa nya-ngu-lta.
 follow-SER-interest **track.ACC** follow-SER-interest see-PST-and.then
 ‘They follow the tracks then find it.’
 Yankunytjatjara (Goddard 1985: 190)

Neither do contextually unique referents have to be marked as such, as can be seen in example (387). Here, the speaker talks about the process of tracking a kangaroo. In this context, the referent of *katji* ‘spear’ in (387b) is unique; it is expressed by a bare noun.

- (387) a. maa-wana-ra malu ngari-nytja-la nyaku-la
 away-follow-SER kangaroo.NOM lie-NMLZ-LOC see-SER
 ‘following it off and seeing where the kangaroo is lying,’
 b. **katji** ila-ra, pungku-la nyuti-ra kulpa-nytja kulpa-nytja
spear.ACC pull-SER hit-SER make.into.bundle-SER return-NMLZ return-NMLZ
 tjuta, palu-nya tjana-nya.
 many.ACC DEF-ACC 3PL-ACC
 ‘pulling out the spear, killing the kangaroo, making it into a bundle, coming back, all those things;’
 Yankunytjatjara (Goddard 1985: 177)

5.4.5 Gooniyandi

Gooniyandi is a Bunaban language spoken in the southern Kimberley region of Western Australia. The language has approximately 100 fluent speakers.⁴² Like Bininj Kun-Wok, Gooniyandi does

⁴²Cf. <http://endangeredlanguages.com/lang/401>.

not have any articles besides the recognitional article *ginharndi*.⁴³ McGregor (1990) does not call the article a recognitional article or marker, although he consistently translates it by ‘you know’. Also its description by McGregor (1990) suggest that *ginharndi* as clearly is a recognitional article.

The determiner *ginharndi*, which is normally glossed ‘you know’ or ‘that one, you know’ by speakers, refers to something which is neither in the immediate speech situation, nor mentioned in the surrounding text. It indicates that, in the speaker’s estimation, the hearer is able to identify the entity being referred to, it being known to him or her. The fact that it is explicitly stated as being shared knowledge is the only clue that is given to the hearer to help him/her identify it. (McGregor 1990: 145-146)

An example of *ginharndi* being used to indicate identifiability based on shared previous experience is shown in example (388). Here, the speaker refers to a shared experience between the speaker and the hearer from the previous day. Note that in this case, more information about the referent is additionally provided in the relative clause following the nominal expression. This is not necessarily the case, though. In (389), only the recognitional article signals that the referent of *booga* ‘the little one’ is identifiable based on shared knowledge between the speaker and the hearer. The utterance is part of a story in which the referent of *booga* ‘the little one’ is a protagonist. Not having been referred to by a nominal expression in 7 clauses, the speaker mentions the referent using a nominal expression together with *ginharndi* to signal that the hearer is supposed to construct the referent as identifiable.⁴⁴

(388) ngooddoo-ngga [**ginharndi** goornboo] wardgilayi-ngangi moolooddja-yidda
 that-ERG ART:RECOG woman she.had.gone-with.us Mulurrja-ALL
 niyaji-ngga barnnginaddi gaddwaroo.
 this-ERG she.returned.me afternoon
 ‘The woman who went to Mulurrja with us brought me back yesterday.’
 Gooniyandi (McGregor 1990: 146)

(389) [booga-ngga **ginharndi-ga**]-waami gindiwa giddari-nga niyaji-yidda.
 baby-ERG ART:RECOG-ERG-INDEF upstream run-3SG this-ALL
 ‘It was the little one, you know, he ran off upstream to this place.’
 Gooniyandi (McGregor 1990: 586)

⁴³McGregor (1990: 144) mentions two anaphoric demonstratives, *niyaji* (proximal) and *niyi* (distal), but their distribution in the examples provided suggest that they are not strictly used for marking anaphoric referents but can also mark deictic ones. Because of that, and because anaphoric referents do not systematically have to be marked as such, I do not include the two anaphoric markers here.

⁴⁴The marker *-waami*, glossed as “indefinite” in example (389) may serve other pragmatic functions rather than indicating indefiniteness. McGregor (1990: 486) mentions that it occurs to mark disjunctions and notes that such “indefinite” markers can express doubt; he also notes that there are too few examples of the marker to be certain about its function.

Example (390) shows the use of *ginharndi* in an establishing context. In this case, the article signals that the hearer should construct the referent as identifiable; the necessary information to do so is then provided in the relative clause following the nominal expression.

- (390) [ginharndi yoowooloo] jijaggiddaa-nhi wambiggoowaari.
 ART:RECOG man we.are.speaking-of.him he.is.going.inside
 ‘The man who we’re talking about is going inside.’
 Gooniyandi (McGregor 1990: 438)

The article *ginharndi* does not occur with deictic referents. McGregor (1990: 144) notes that Gooniyandi has two spatial demonstratives, proximal *ngirdaji* and distal *ngooddoo*. An example is given in (391).

- (391) [ngooddoo yaanya yaanya balma] milawa niyaji boolooboowa.
 DEM:DIST other other fork you.will.see.it DEM:PROX you.will.follow.it
 ‘That other (road–not the one you are following now); you see the other road, well follow it.’
 Gooniyandi (McGregor 1990: 455)

Examples (392) shows that an anaphoric referent is not marked by *ginharndi* either. The anaphoric referent of *liddi* ‘guts’ is expressed by a bare noun. The same is shown in (393) for a contextually unique referent. Both examples are part of a story about a hunting trip. Having caught three crocodiles, they describe the part of the preparation and cooking. The referent of *mirlo* ‘liver’ mentioned in (393) is contextually unique and expressed by a bare noun as well.

- (392) a. **liddi** ngab-a-yiddi.
 guts eat-PROG.NOM.1
 ‘We were eating the guts.’
 b. woobi-yidda:: liddi ngab-jidda.
 cook-NOM.1 guts eat-NOM.1
 ‘We cooked the guts and ate them.’
 Gooniyandi (McGregor 1990: 575-576)
- (393) **mirlo**-nyooloo doow-wani-widda-yi:: wili.
 liver-etc get-it-NOM.3PL-DU finish
 ‘They got out the liver and so on.’
 Gooniyandi (McGregor 1990: 575)

5.4.6 Recognitional articles: summary and outlook

This section presented a number of recognitional articles, whose crosslinguistic occurrence seems concentrated in the areas of Australia and Papunesia. We saw that all five recognitional articles

presented are used to mark a referent that is unambiguously identifiable by the speaker and the hearer based on shared knowledge.

In the two Papuan languages Oksapmin and Lavukaleve, we also saw that the recognitional article is not only used to refer to personal knowledge shared between the speaker and the hearer but to culture-specific knowledge shared by the entire speech community. As was mentioned in Section 3.2.3, this use could simply serve to create a higher degree of proximity and a sense of community between speakers, as was already noted in Diessel (1999: 106). Especially the recognitional articles of the two Papuan languages Oksapmin and Lavukaleve discussed here belong to speech communities where a strict distinction between general world knowledge and shared personal knowledge may not be applicable in a way it would be to the speech communities of official national languages. The earlier linguistic descriptions of recognitional uses were based on official national languages with large speech communities. The speech communities of both Oksapmin and Lavukaleve are rather small with approximately 8000 and 1700 speakers each, and both languages are spoken in rather remote areas with rather tight-knit communities. Of course, a much more principled investigation would be required, but the data presented here strongly suggests that a simple division into general world knowledge as opposed to shared personal knowledge may not be equally applicable to speech communities across languages. We observe a similar situation in the three Australian languages presented in this section. Bininj Kun-Wok, Yankunytjatjara, and Gooniyandi have rather small speech communities as well, at least in comparison to modern national languages, but not necessarily in the context of Australian Indigenous languages.

In addition, it was mentioned for Yankunytjatjara that cultural factors such as taboos and tendencies to avoid using names may favor the use and development of referential markers that indicate shared knowledge about a referent so that a lexically more general term can be used and too direct reference can be avoided. All these observations strongly suggest that further research is necessary, using a more fine-grained distinction of different types of (shared) knowledge and experience, taking into account socio-linguistic factors.

Moreover, the patterns discussed in this section clearly fit into the wider picture of socio-linguistic factors influencing grammars as shown in Lupyan & Dale (2010, 2016), Perkins (1992), Trudgill (2008, 2011), Wray & Grace (2007), amongst others. It seems especially related to the observation that complex pronominal or deictic systems tend to be found in small and close-knit speech communities first noted by Perkins (1992) and various studies since then (e.g. Dench 1994b, Tadmor 2015, Trudgill 2015). Including the data on recognitional articles presented here could extend such observations to systems with very specialized referential markers, which of course will require more detailed research in the future to allow a more conclusive picture of the crosslinguistic distribution of recognitional markers.

5.5 Summary

This chapter presented different examples of the four types of articles in the definite domain: definite, anaphoric, weak definite, and recognitional articles. Their occurrence across different types of definite referents is summarised in Table 5.12.

For definite articles, Sections 5.1.3, 5.1.4, and 5.1.5 showed that the occurrence of definite articles with deictic and absolutely unique referents as well as with proper nouns is possible in some languages while definite articles are blocked in such contexts in others. I argued that the compatibility of definite articles with deictic referents can be accounted for by the origin of the definite article, which goes back to a spatial demonstrative in most cases. In the case of Kaqchikel, where this is not the case, we saw that the definite article is incompatible with deictic referents. Similarly, we saw that many definite articles do not mark absolutely unique referents. However, definite articles can be used in such contexts in some languages, and as was briefly mentioned, their extension to absolutely unique referents can be explained by their function of marking such referents as prominent in the discourse.

Anaphoric articles correspond to the type of articles labeled strong definite articles in Schwarz (2009). Such articles are indeed not uncommon in the world's languages. Their core function being the expression of anaphoric referents, they do not mark deictic referents and they may differ as to which other pragmatic definite referent types they can encode. In Table 5.12, Komnzo is marked as compatible with a deictic referent because it can co-occur with a spatial deictic demonstratives in such contexts. Section 5.2.4 argued that anaphoric articles may generally be able to encode establishing and recognitional referents, providing examples from various languages. Future work will be needed to determine to what extent such uses are systematic across languages.

Section 5.3 gave an overview of weak definite articles, the counterpart of anaphoric articles. I described its use with data from Fering and German, which are the two showcases of weak definite articles. The following Sections discussed data from Hausa, Lakota, and Ma'di, which were suggested as examples of languages with weak definite articles in the literature. In all four cases, I argued against an analysis of the relevant marker as a weak definite article. For Hausa, Section 5.3.2 showed that the potential weak definite article is in fact an anaphoric article similar to the anaphoric article in Akan. The potential strong definite article, on the other hand, is an additional anaphoric demonstrative marker that can be used in certain anaphoric cases. Factors to influence the variation between those two markers may be the distance to the antecedent, discourse salience and recoverability of the referent. Section 5.3.3 then showed a similar situation in Lakota. I argued that the potential weak definite article is in fact a definite article which is generally also used in anaphoric contexts. Lakota has an additional marker that can be used in certain anaphoric contexts depending on the topicality and discourse salience of the referent.

Urama, yet another language with what may look like a weak definite and an anaphoric article was discussed in Section 5.3.4. I showed that also in this language, the presence of an anaphoric article, which is used systematically in anaphoric contexts, does not entirely block the use of the definite article in at least certain anaphoric contexts. The only other language proposed to have a weak definite article is Ma'di. Section 5.3.5 showed that an analysis of the marker in question as a weak definite article is not tenable in this case either. I argued that the presumed weak definite article should not be regarded as an article in the first place but would better be described as a clausal determiner or backgrounding marker.

Table 5.12: The distribution of definite, anaphoric, and recognitional articles

article	R _{deictic}	R _{anaph}	R _{recog}	R _{establ}	R _{rel-bridge}	R _{sit.unique}	R _{cont.unique}	R _{u-bridge}
definite								
Kaqchikel (<i>ri</i>)	X	✓	✓	✓	✓	✓	✓	✓
Mokpe (<i>è/H</i>)	(✓)	✓	✓	✓	✓	✓	✓	✓
anaphoric								
Limbun (<i>fō</i>)	X	✓	(✓)	X	X	X	X	X
Komnzo (<i>ane</i>)	(✓)	✓	X	✓		X	X	
Akan (<i>no</i>)	X	✓	✓	✓	✓	(✓)	X	X
weak definite								
Fering (<i>a, ...</i>)	X	X			X	✓	✓	✓
German (<i>zum, vom, ...</i>)	X	X	X	X	X	✓	✓	✓
recognitional								
Oksapmin (<i>max</i>)	X	X	✓	(✓)				X
Lavukaleve (<i>mea</i>)	X	X	✓	(✓)				X
Bininj Kun-Wok (<i>nawu</i>)	X	X	✓	✓				X
Yankunytjatjara (<i>panya</i>)	X	X	✓	(✓)				X
Gooniyandi (<i>girnharndi</i>)	X	X	✓	(✓)				X

Section 5.4 discussed recognitional articles, which are crosslinguistically very rare, as well. The 5 examples from the sample all came from Papuan and Australian languages, which suggests

that the presence of recognitional articles may be an areal feature. However, it is also clear that the recognitional function is not necessarily well described in the sources so that the areal bias could reflect a bias in language documentation. Contrary to what earlier work on recognitional markers suggested, recognitional articles are often also used to make reference to culture-specific concepts. I argued that this may reflect different cultural factors in communities of intimates (that the languages with recognitional articles all are) vs. communities of strangers, fitting in with a line of research that has observed this for deitic and demonstrative systems in a more general way. Thus, more language-specific descriptions of such contexts as well as theoretical work is needed to arrive at a better description of the properties and the crosslinguistic distribution of recognitional articles.

Chapter 6

Articles in the indefinite domain

This chapter provides an overview of the articles from the indefinite domain, discussing examples of indefinite (Section 6.1), presentational (Section 6.2), exclusive-specific (Section 6.3), and nonspecific articles (Section 6.4). In addition to the discussion of examples, this chapter makes two important general observations concerning articles in the indefinite domain. The first one concerns a number of nonspecific articles (or markers) in North America. In Section 6.4.4, I show that some nonspecific markers can be traced back to verbal irrealis markers. The second observation concerns a type of indefinite articles that I call presentational articles. Presentational articles only occur with discourse-prominent specific and nonspecific referents. This is important for the diachrony of indefinite articles insofar, as it shows that their diachronic semantic extension from specific to nonspecific contexts and their pragmatic extension from prominent to less prominent referents are independent from each other (Section 6.2.5).

In order to distinguish different types of articles in the indefinite domain, the relevant contexts are constructions in which the referent is expressed as an argument within the scope of negation, a question, an intensional operator or certain types of modal operators. The following examples briefly shows such contexts with data from English, since many of the examples in this chapter are variations of these contexts.

Example (394) shows negation scoping over the argument with the referent in question. Many of the examples in this chapter are negated existential constructions, simply because these are the commonly used examples with negation in the grammars. In such negation contexts, the referent is nonspecific by default. The same holds for referents within the scope of a question as in example (395) and the referents in (396), which are within the scope of a modal operator, modal being used in a very wide sense here. Therefore, only indefinite and nonspecific as opposed to exclusive-specific articles are expected to occur in such contexts. Note that inclusive-specific articles are expected to be felicitous in these contexts as well, since they also allow for a definite interpretation

of the referent. Importantly, this is the only interpretation allowed in such contexts, given that the referent cannot be specific.

(394) Negation

- a. *I do not see **a(ny) fish** in the tank.*
- b. *I do not have **a(ny) pet**.*
- c. *There is **no bakery** at the corner.*

(395) Question

- a. *Do you see **a(ny) fish** in the tank?*
- b. *Do you have **a pet**?*
- c. *Is there **a bakery** at the corner?*

(396) Modals

- a. *Please, get me **some tomatoes** when you go to the store.*
- b. *Can you bring me **some apples** from the kitchen?*
- c. *If I had bought **a tent**, I would have joined you on the trip.*

In example (397), the referents are within the scope of an intensional predicate, allowing for both specific and nonspecific interpretations. In languages with indefinite articles, the interpretation of the referent remains ambiguous, unless other grammatical markers or contextual cues resolve the ambiguity. In languages with an exclusive-specific article, only the specific interpretation should be felicitous; vice versa, nonspecific articles are expected to only allow for a nonspecific interpretation of the referent.

(397) Intensional predicate

- a. *I want see **a movie**.*
- b. *I wish I had **a cat**.*
- c. *I am looking for **a bike**.*
- d. *I hope to find **a present** for my friend.*

6.1 Indefinite articles

Indefinite articles encode specific and nonspecific referents. In this Section, I present three examples of this article type: the indefinite articles in Tz'utujil (Section section 6.1.1), Sri Lanka Malay (Section section 6.1.2), and Bonan (Section section 6.1.3).

- b. ximba p'=k'olbek.
 go.1SG to=**wedding**
 'I went to the wedding.'
 Tz'utujil (primary data)

In Tz'utujil, the marking of plural in contrast to unmarked singular is commonly found with human referents, but not obligatory with most other types of nouns (Dayley 1985: 140f). In (402) and (403), we see that the indefinite article can be used with plural nouns as well. Note that in (403), the indefinite article can be omitted; however, the referent of *wixb'ij* 'friends' would then receive a definite interpretation.

- (402) a. atrejben [***(jun)** xtan] chu jay.
 wait.s:3SG.O:2SG **ART:INDEF** woman outside house
 'A lady is waiting for you outside.'
- b. atkiyben [***(julee)** xtan-ij] chu jay.
 wait.s:3PL.O:2SG **ART:INDEF-PL** woman.PL outside house
 'Some ladies are waiting for you outside.'
- Tz'utujil (primary data)

- (403) [**julee** wixb'-ij] xinkeresta' chaq'a' iwir.
ART:INDEF.PL friend-PL came.see night yesterday
 'Some friends came to visit last night.'
- Tz'utujil (primary data)

The article *jun* is an indefinite article because it is also used to encode nonspecific referents. Thus, it does not disambiguate between a specific and nonspecific interpretation of the referent. Examples (404), (405), and (406) illustrate its use with nonspecific referents.

- (404) k'ool [***(jun)** atz'ib'ab'el]?
 EXIST **ART:INDEF** pen
 'Do you have a (any) pen?'
 Tz'utujil (primary data)
- (405) nk'atzin chwa' [***(jun)** atz'aqoneel].
 need to.1SG **ART:INDEF** architect
 'I need an (any) architect.'
 Tz'utujil (primary data)
- (406) nwajo' [***(jun)** ch'eech'].
 want.PRES.S:1SG.O.3SG **ART:INDEF** car
 'I want a (any) car.'
 Tz'utujil (primary data)

Even though examples (404) to (406) show that the indefinite article *jun* is used to mark nonspecific indefinite referents, it appears to be less natural in at least certain nonspecific contexts involving the existential marker *k'ool*.² In the context of a question as in (407) and (408a), the use of *jun* is judged as not obligatory by my consultants. In the answer in (408b), on the other hand, *jun* is required like in the other indefinite contexts seen above.

(407) k'ool [(jun) k'olb'al] ab'aar nkatwawa' chpaam jawa tinamit?
 EXIST ART:INDEF place where can.eat inside DEM:PROX village
 'Is there a place where one can eat (in a restaurant) in this village?'
 Tz'utujul (primary data)

(408) a. k'ool [(?jun) tiyoxajaay] chpaam jawa tinamit?
 EXIST ART:INDEF church inside DEM:PROX village
 'Is there a church in this village?'
 b. jee' k'ool *(jun) tiyoxajaay.
 yes EXIST ART:INDEF church
 'Yes, there is a church.'
 Tz'utujil (primary data)

Despite the fact that *jun* shows exceptions to its use with the copula *k'ool*, its use in specific and nonspecific contexts is systematic. Moreover, examples (64) and (65) in Section 2.2.4 showed that *jun* does not occur with anaphoric or contextually unique referents. Therefore, Tz'utujil *jun* needs to be analysed as an indefinite article.

6.1.2 Sri Lanka Malay

Sri Lanka Malay is a Malayo-Polynesian language spoken in Sri Lanka. It substantially differs from Standard Malay and should not be taken as a dialect; according to Nordhoff (2009: 3), the two languages are mutually unintelligible. Sri Lanka Malay draws much of its lexicon from Standard Malay, while its grammatical structure is strongly influenced by its contact languages Sinhala and Tamil, amongst others (Nordhoff 2009: 3). According to the Endangered Languages Project, there are approximately 46,000 speakers of Sri Lanka Malay.³

Sri Lanka Malay does not have articles in the definite domain; it only uses the indefinite article presented in this Section. Nordhoff (2009: 319) mentions that the Sri Lanka Malay indefinite article *hatthu* can be realized as *(h)at(t)hu*; it has a distinct form from the numeral 'one' *sat(t)hu*. The article is a clitic and can precede or follow the noun, or it can occur as an enclitic and a proclitic on both sides of the noun at the same time. Regarding its flexible position, Nordhoff (2009: 319-320) notes that "[i]t is probable that the preference for one or the other is idiolectal, but speakers

²Neither Dayley (1985) nor Tz'utujil Tinaamitaal (2007) discuss this.

³Cf. <http://www.endangeredlanguages.com/lang/3568>.

happily accept any of the two versions when asked”; he also notes that double marking often takes place with loan words, which may go back to a parallel Sinhala pattern.

In (409), we see the indefinite article *hatthu* in a typical presentational context. The referent of *aanak raaja* ‘prince’ is introduced using the indefinite article in (409a). When the referent is mentioned again two clauses later, it is identifiable and expressed as a bare noun.

- (409) a. Derang anà-baalek sajja=jo, sithu=ka panthas [**hatthu** Aanak raaja]
 3PL PST-turn only=EMPH there=LOC beautiful ART:INDEF king child
 su-aada! ...
 PST-exist ...
 ‘Just when they turned around, there was a beautiful prince! ...’
- b. **Aanak raaja** su-biilang: ...
child king PST-say ...
 ‘The prince said: ...’
- Sri Lanka Malay (Nordhoff 2009: 678)

Other examples in which *hatthu* marks a referent as non-identifiable but specific are given in (410) to (413). In (410) and (411), the indefinite article precedes the noun. In (412), it follows it, and we can see it in both positions in (413).

- (410) Sithu=ka, [**hatthu** bissar beecek caaya Buruan] su-duuduk.
 there=LOC ART:INDEF big mud colour bear PST-exist
 ‘There was a big brown bear there.’
 (Nordhoff 2009: 168)
- (411) mareng dovulu incayang=nang HSBC=ka [**hatthu** pukuran] se-daapath.
 yesterday earlier 3SG.polite=DAT HSBC=LOC ART:INDEF job PST-get
 ‘he got a job at the HSBC bank the day before yesterday.’
 Sri Lanka Malay (Nordhoff 2009: 682)
- (412) [Bissar ruuma **hatthu**] aada.
 big house ART:INDEF exist
 ‘There is a big house.’
 Sri Lanka Malay (Nordhoff 2009: 212)
- (413) Itthu haari=ka=jo anak pompang duuva=nang [...] [**hatthu** kiccil jillek aajuth
 DIST day=LOC=EMPH child girl two=DAT [...] ART:INDEF small ugly dwarf
hatthu=yang] su-kuthumung.
 ART:INDEF=ACC PST-see
 ‘On that very day, the two girls saw a small ugly dwarf.’
 Sri Lanka Malay (Nordhoff 2009: 685)

Nordhoff (2009: 198) mentions that *hatthu* is also required if nouns are used predicatively as in (414).

- (414) Farook [***(atthu)** maaling].
 Farook ART:INDEF thief
 Farook is a thief.
 Sri Lanka Malay (Nordhoff 2009: 198)

In addition to marking a referent as non-identifiable and specific, *hatthu* also occurs systematically with nonspecific referents, which warrants its status as indefinite article. In (415), *kaar* ‘car’ is the direct object of an intensional predicate; its non-identifiable referent is therefore ambiguous between a specific and a nonspecific interpretation. We see that it is marked by *hatthu*. Example (416) shows the indefinite article within the scope of negation. In the presence of the negative marker *thraa*, it encodes the referent *pada* ‘child’ as nonspecific.

- (415) Se=dang se=ppe pukujan pada=nang [baaru **hatthu** kaar] maau.
 1SG=DAT 1SG=POSS work PL=DAT new ART:INDEF car need
 ‘I need a new car for my work.’
 Sri Lanka Malay (Nordhoff 2009: 651)

- (416) Ini aanak pada=nang [time **hatthu** thraa].
 PROX child PL=DAT time ART:INDEF NEG
 ‘These children do not have any time.’
 Sri Lanka Malay (Nordhoff 2009: 257)

In examples (417) and (418), the referents marked by *hatthu* are part of a non-episodic events. Therefore, the referents are nonspecific in these examples as well.

- (417) [**Hathu** oorang] kala-pasiith, [**hathu** chance] masà-kaasi ithu oorang=nang=le.
 ART:INDEF man if-trouble ART:INDEF chance must-give DIST man=DAT=ADDIT
 ‘When a person has trouble, we must give him a chance.’
 Sri Lanka Malay (Nordhoff 2009: 654)

- (418) Kithang=nang [**hathu** application] mà-sign kamauvan vakthu=nang=jo, kithang
 1PL=DAT ART:INDEF application INF-sign want time=DAT=EMPH 1PL
 arà-pii inni politicians pada dikkath=nang.
 NPST-go PROX politicians PL vicinity=DAT
 ‘When we want to sign an application, we approach these politicians.’
 (Nordhoff 2009: 251)

The indefinite article *hatthu* can also be combined with a negative polarity item. This is shown in (419) and (420), where *hatthu* co-occurs with *on* ‘any’.

(419) Kithang **hatthu=oorang=pon** thàrà-iingath.
 1PL ART:INDEF=man=any NEG.PST-think
 ‘We cannot think of any person’
 Sri Lanka Malay (Nordhoff 2009: 673)

(420) See pukaran=**hatthu=pon** thama=gijja, ruuma=ka arà-duuduk.
 1SG work=ART:INDEF=any NEG.NPST=make house=LOC NPST-stay
 ‘I don’t do any work, I stay at home.’
 Sri Lanka Malay (Nordhoff 2009: 673)

Having established that *hatthu* is indeed an indefinite article for marking referents as specific or nonspecific in a systematic way, the following example shows that it does not occur with a definite referent. Like the anaphoric referent shown in example (409b) in the beginning of this section, the contextually unique referents in (421) are expressed as bare nouns.

(421) **Oorang mlaayu** siithu=dering dhaathang=apa **cinggala raaja=nang=le** anà-banthu.
 man Malay there=ABL come=after Sinhala king=DAT=ADDIT PST-help
 ‘The Malays came from there and helped the Sinhala king.’
 Sri Lanka Malay (Nordhoff 2009: 613)

Thus, we must conclude that *hatthu* in Sri Lanka Malay is indeed an indefinite article.

6.1.3 Bonan

Bonan (or Bao’an Tu) is a Mongolic language spoken in the Gansu and Qinghai Provinces in China (Fried 2010: 1). According to Fried (2010: 6), Bonan has approximately 4000 speakers. Bonan only has an indefinite article presented in this Section; it does not make use of article in the definite domain.

Example (422) shows that the indefinite article *-gə* occurs with the noun *ahku* ‘Buddhist monk’ to introduce it as a non-identifiable referent:

(422) hkutə tɕæci silaŋ=da o=ku tɕhəχaŋnaŋ atɕaŋ pantɕhe tɕə
 yesterday Jiashi Xining=LOC go=IPFV.NMLZ time 3SG bus on
 ahku=gə tɕolaxa-tɕə.
 Buddhist.monk=ART:INDEF meet-PFV
 ‘Jiashi met a monk on the bus when he went to Xining yesterday.’
 Bonan (Fried 2010: 82)

We see the marker *-gə* with other specific referents in examples (423) to (426). In (423) and (424), the article occurs in the object position of a transitive clause. Example (425) shows that the indefinite article *-gə* is also used in the subject position.

(423) atɕaŋ dzəjaŋ=gə=tɕə khidzə=gə ap-tɕo.
 3SG bowl=ART:INDEF=and chopsticks=ART:INDEF buy-IPFV.O
 ‘He bought a bowl and a pair of chopsticks.’
 Bonan (Fried 2010: 167)

(424) atɕaŋ mənda [mehtox tɕəχaŋ=gə] oχ-tɕo.
 3SG 1SG.LOC flower white=ART:INDEF give-IPFV.O
 ‘He gave me a white flower.’
 Bonan (Fried 2010: 146)

(425) [deda ŋarwaŋ=gə] pətə χeɕaŋ pitəβaŋ khər-saŋ=nə
 grandfather powerful=ART:INDEF 1PL.EX.COLL Leru small be.required-PFV.NMLZ=ACC
 thorun hko=nə khərə=tɕə khəl-saŋ.
 head big=ACC be.required=QUOT.IMPF say-POS
 ‘A powerful grandfather said, “as we have to have the small Leru festival, [we] must have a big well”.’
 Bonan (Fried 2010: 118)

(426) dedə=la laχci tɕo-tɕə sarə=la=nə χamdə hkuraŋ=gə jiχ-sa
 grandfather=PL towel wear-IPFV young=PL=ACC together circle=ART:INDEF do-COND
 thəke-saŋ.
 permit-EPIST
 ‘The grandfathers can wear a towel and (dance) in a circle with the young (men).’
 Bonan (Fried 2010: 200)

Fried (2010: 47) notes that indefiniteness in object positions is signalled by the absence of the accusative marker *-nə* so that the indefinite article is not required in this position. A number of specific and nonspecific referents in object positions are indeed expressed as bare nouns in the examples provided by Fried (2010). One such example is given in (427) below.

(427) lapsi=da saŋ χərə-tɕə tɕhəma=nə χara-tɕo.
 Lapsi=LOC offering burn-IPFV Chuma=ACC taunt-O.IPFV
 ‘And at Lapsi (they) burn offerings and taunt Chuma.’
 Bonan (Fried 2010: 329)

Nevertheless, the previous sentences in (423) and (424) attest that the indefinite article does occur in the object position. Example (428) not only shows that *=gə* can be used with objects but shows that it is compatible with the accusative marker as well.

- (428) nadamada khəl-də=ku tɕəχaŋnaŋ jimaŋ=gə=nə
 long.ago speak-IMP=IPFV.NMLZ time goat=ART:INDEF=ACC
 sugə-saŋ=tɕə khəl-na.
 argue-EPIST=QUOT.IPFV speak-DUR
 ‘It is said that long ago, when (we) were fighting, (we) were arguing over a goat.’
 Bonan (Fried 2010: 330)

As an indefinite article, *-gə* must also occur with nonspecific indefinite referents. Examples of contexts with referents whose interpretation is ambiguous between being specific and nonspecific are given in (429) and (430). They show that a noun marked by *-gə* is compatible with the nonspecific interpretation of the referent as well.

- (429) dzoma [htɕəχta χənə=gə] ap=ku taraŋ əŋ-tɕo.
 Droma bicycle new=ART:INDEF take=IPFV.NMLZ desire come-IPFV.O
 ‘Droma wants to buy a new bike.’
 Bonan (Fried 2010: 82)

- (430) pə dzoma=da samtəχ=gə oχ=ku taraŋ əŋ-na
 1SG Droma=LOC gift=ART:INDEF give=IPFV.NMLZ desire come-DUR
 ‘I want to give a gift to Droma.’
 literally: ‘The desire that I give a gift to Droma comes.’
 Bonan (Fried 2010: 282)

That the article *-gə* does not occur in definite contexts is shown in examples (431a), (431b), and (432) below. Definite referents in Bonan are expressed by bare nouns. In (431a), *səməla* ‘the boxes’ are contextually unique in the context of the story told, the same holds for *gərgəŋ* ‘the teachers’ in (432) in a story about a certain school. The sentence in (431b) features the anaphoric referent of *ula* ‘mountain’. All three definite referents are expressed as a bare nouns.

- (431) a. **səme**=la=nə **ula**=da dɒbla-tɕə o khəŋ-na
god.box=PL=ACC **mountain**=LOC raise-IPFV go be.required-DUR
 ‘(We) must carry the god boxes to the mountain.’
 Bonan (Fried 2010: 327)
- b. de χara-tɕə=ku tɕənsan sowu **ula**=nə oχ-to=tɕi-saŋ
 just taunt-IPFV=NMLZ reason important **mountain**=ACC give-PFV=S.QUOT.IPFV-EPIST
 ‘But we consider the main reason for taunting (them) to be that we gave them the mountain.’
 Bonan (Fried 2010: 331)

- (432) *gərgən=da cəkə sokərgə-təə*
teacher=LOC very respec-PFV
'Be very respectful toward the teachers.'
Bonan (Fried 2010: 320)

Since *-gə* is systematically used with specific and nonspecific referents while being absent in definite contexts, it is yet another example of an indefinite article.

6.2 Presentational articles

In a number of languages, we find markers that systematically occur with specific referents in discourse-prominent positions. The presence of such markers exceeds the occasional use of the numeral 'one' in this type of contexts (cf. Section section 4.3). Nevertheless, this type of article is somewhat different from the other indefinite articles as seen in the previous sections, as it mainly occurs with discourse-prominent referents. I understand discourse prominence in the sense of von Heusinger & Schumacher (2019).⁴ Their characterization of discourse prominence is given in (433). Discourse-prominent referents thus stand in relation with less prominent referents. What I mean by discourse-prominent referents regarding the use of presentational articles are those referents that are newly introduced in the discourse, that correspond to the center of attention of the current discourse segment and often to the initial element in a topic chain.

- (433) Characterization of prominence for discourse pragmatics and as a structure building principle for discourse representation
- a. Def.1: Prominence is a relational property that singles out one element from a set of elements of equal type and structure.
 - b. Def.2: Prominence status shifts in time (as discourse unfolds).
 - c. Def.3: Prominent elements are structural attractors, i.e. they serve as anchors for the larger structures they are constituents of, and they may license more operations than their competitors.
- (von Heusinger & Schumacher 2019: 119)

Presentational articles often appear to be at an intermediate stage from the numeral 'one' with occasional presentational functions on one end of a scale and with the indefinite article on the other end. This corresponds to the traditional grammaticalization path of indefinite articles, which was

⁴Besides formulating the criteria for discourse prominence, von Heusinger & Schumacher (2019) offer an overview of related relevant notions and approaches such as referential activation (Chafe 1976, Lambrecht 1994), Centering Theory (Grosz et al. 1995), accessibility (Ariel 1990), givenness (Gundel et al. 1993), and saliency (Chiaros et al. 2011).

presented in Section section 4.3. Therefore, it seems plausible to analyse presentational articles as emerging specific or indefinite articles. I will argue in the present section that this is not the case.

A presentational article is not necessarily a marker at an intermediate or “unstable” stage between the numeral ‘one’ and the indefinite article. There are two reasons for this. First, the use of such markers is systematic, even though it is restricted to discourse-prominent referents from the indefinite domain. Second, their use is not necessarily strictly bound to specific referents. Data from several languages discussed here suggests that presentational articles may also be able to encode nonspecific referents. Therefore, I treat such markers as a type of indefinite articles, since both are used to encode specific and nonspecific referents. The only difference between indefinite and presentational articles is that the use of the latter is constrained by additional pragmatic factors.

This section shows examples of presentational articles from Lango (Section 6.2.1), Bilua (Section 6.2.2), Teotepec Chatino (Section 6.2.3), and Kashibo-Kakataibo (Section 6.2.4). Section 6.2.5 then reassesses the development from the numeral ‘one’ to the indefinite article in the light of the data presented.

6.2.1 Lango

Lango is a Western Nilotic language spoken in central Uganda. Noonan (1992: 1) estimates that the language has between 300,000 and 800,000 speakers. Lango does not have any articles other than the presentational article described in this section.

The presentational article has the form *-mórô* (the initial nasal is phonologically assimilated to the last segment of the noun that the article occurs with). It systematically marks referents as specific, but it only occurs in certain types of contexts. Examples (434) and (435) below illustrate the use of *-mórô*.

- (434) *twòl-lórô òkàò àtín.*
 snake-ART:PRES bite.PERF.3SG child
 ‘A snake bit the child.’
 Lango (Noonan 1992: 162)

- (435) *pónó'-mórô tíê ì nê òt.*
 pig-ART:PRES be.present.HAB.3SG in back house
 ‘There’s a pig behind the house.’
 Lango (Noonan 1992: 162)

However, specific referents do not generally need to be marked by *-mórô*, as the following two examples show. Even though the referents of *lócà* ‘man’ in (436) and *mé* ‘girl’ (437) are not identifiable by the hearer, they do not require a specific marker.

(436) ànjéô lócà à têk à ryèk.
 know.HAB.1SG man ATTR strong ATTR wise
 ‘I know a strong and wise man.’
 Lango (Noonan 1992: 164)

(437) nwàŋ gìn òlárò pákô mé jòm.
 long.ago they compete.PFV.3PL girl for marriage
 ‘Long ago they were competing to marry a girl.’
 Lango (Noonan 1992: 294)

Noonan (1992: 147, 162) notes that indefinite referents, expressed as subjects in existential sentences, are often accompanied by *-mórô*, and describes the distribution of *-mórô* in indefinite contexts as follows:

Not all indefinites are marked with *-mórô*; in fact, the majority are not [...] [It] is almost always found in subject position. The reason for this is that the Lango sentence is organized roughly on an old-new continuum. Subjects are always interpreted as definite unless specifically made indefinite with *-mórô* (hence the association with subjects in existential constructions). In other sentence positions, NPs are assigned definite or indefinite interpretations pragmatically, and do not require an explicit indefinitizer like subjects. (Noonan 1992: 162)

Moreover, Noonan (1992: 162) also notes that the use of *mórô* is the only way a subject can be marked as indefinite. This syntactic restriction or condition on the presentational article is the result of the interaction between information structure, referentiality, and word order in Lango. The preverbal position is reserved for topics; by default, the subject is realized in this position and is interpreted as definite (Noonan 1992: 119). Because of this strong connection between the subject and topicality and by extension definiteness, it is not surprising that the presentational article in Lango is also strongly associated with subjects and the clause-initial position. Since the identifiability of the referent is less strongly associated with other syntactic functions and positions, one could imagine that there is no functional pressure for *-mórô* to be extended to other contexts. However, the fact that *-mórô* is used in syntactically restricted contexts and may not be very frequent in general does not make the article optional. Because the distribution of *-mórô* can be described in this way, it is systematic, making *-mórô* a presentational article.

While Noonan (1992) does not further distinguish between specific and nonspecific referents in the indefinite domain, we do not only find the article *-mórô* with specific referents as was the

case in the previous examples. Example (438) shows the presentational article in a nonspecific context:

- (438) cèn-nórô pé
ghost-ART:PRES be.absent.HAB.3SG
‘There are no ghosts.’
(Noonan 1992: 147)

This use is evidence for the necessity to separate the pragmatic extension of emerging indefinite articles (from discourse-prominent or topical to less prominent or non-topical referents) from their semantic extension (from specific to nonspecific contexts). The presentational article in Lango shows that the articles in the indefinite domain can be used for both specific and nonspecific referents, while being strongly restricted in their distribution based on information-structural and syntactic factors. The presentational article in Lango also suggests that there is no inherent reason to assume that such a distribution would be “unstable” in some way and necessarily lead to the development of an indefinite article that is used with specific and nonspecific referents regardless of their discourse prominence.

6.2.2 Bilua

Bilua belongs to the group of Papuan languages and is spoken on the Vella Lavella island of the Solomon Islands; Obata (2003: 1) estimates the number of speakers between 8000 and 9000. Bilua does not have other articles than the presentational one described here.

As examples (439) and (440) show, the marker *kama/kala* (feminine, masculine forms) occurs with specific referents when they are introduced in the discourse and when they are not yet identifiable by the hearer in (439a) and (440a). In addition, both specific referents *niabara* ‘canoe’ and *bazu-bazulao* ‘folktale’ stay the center of attention in the immediately following utterances shown in (439b) and (440b).

- (439) a. Se ta ke ere=v=e [kala niabara].
3PL TOP 3PL make=3SG.M.O=REMPST ART:PRES.M.SG war.canoe
‘They made a war canoe.’
b. Ne=a niabara ta a=daite Bobe=vo.
PROX.SG.M=LNK war.canoe TOP 1SG=grand.parent Bobe=3SG.M
‘This war canoe was my grandfather Bobe’s.’
Bilua (Obata 2003: 79)

- (440) a. Anga ta aqa zari=a vai bazu-bazut=o [kala
 1SG TOP 1SG.O:3SG.F want=PRES O:3SG.M RED-tell=NOM ART:PRES.SG.M
 bazu-bazulao], ...
 RED-folktale ...
 ‘I want to tell a folktale, ...’
- b. ... ni komi=a bazu-bazulao ta pui matu tuvevo=a=ma, ...
 ... and PROX.SG.F=LNK RED-folktale TOP NEG very true=LNK=3SG.F ...
 ‘... and this folktale, it is not very true, ...’
- c. ...melai silo-silo=a=mu kemai ibue=k=o
 ...but RED-small=LNK=3PL S:3PL.O:3PL make.quiet=O:3SG.F=NEARFUT
 k=i=ke=ve=ma.
 O:3SG.F=say=3PL=REMPST=3SG.F
 ‘...but people make children quiet by telling this story.’
 Bilua (Obata 2003: 287-288)

Like in Lango, the presentational article in Bilua is also attested with nonspecific referents. This is shown in examples (441) and (442) below.

- (441) ...enge ta nge=qai zari=a tu k=ov=o [kama
 ...1PL.EXCL TOP 1PL.EXCL=O:3SG.F want=PRES IRR O:3SG.F=get=NOM ART:PRES.SG.F
 uri=a=ma saev=o] ...
 good=LNK=3SG.F survive=NOM
 ‘...we want to get a good life ...’
 Bilua (Obata 2003: 93)

- (442) Ko=lupao=va, ti ko=noqoe=k=ou ko ore ma tataikili ma esa
 3SG.F=dislike=PRES and.then S:3SG.F=hold=O:3SG.F=FUT 3SG.F tree or trunk or maybe
 [kama pata-pata].
 ART:PRES.SG.F RED-stump
 ‘If she dislikes it, she will hold a tree, a trunk, or maybe a stump.’
 Bilua (Obata 2003: 129)

The following two examples show that the presentational article in Bilua is indeed restricted to discourse-prominent referents; less topical referents, like *juli* ‘banana tree’ and *tuto* ‘coconut shell container’ in (443) and (444), are expressed as a bare noun and do not receive an article.

- (443) ...vo ta o=marong=a juli=ko raki ale.
 ...3SG.M TOP 3SG.M=sleep=PRES banana.tree=3SG.F root in
 ‘...[he arrived at the garden, and then] he slept at the base of a banana tree.’
 Bilua (Obata 2003: 290)

- (449) [ska ni nʔi ti^C-kwa^F], [ska yu na^I Mɔ^B]
 ART:PRES 3SG HAB.to.be just-there ART:PRES man PROG.name Filemon
 ‘a respected person who lives up there, the man named Filemon’
 Chatino (McIntosh 2015: 251)

That *skaʳ* does not occur with all types of specific referents is shown in the following two examples. With specific referents that are not relevant to the discourse or whose identifiability does not play a role in the current discourse situation, the presentational article *skaʳ* is not used. Examples of this type of referents, expressed by bare nouns, are given in (450) and (451).

- (450) jyku^r nda³¹nye^r nsna²³.
 eat.COMPL Daniel apple
 ‘Daniel ate an apple.’
 Chatino (McIntosh 2011: 94)

- (451) mt7o³¹-(n)skwa⁰⁽³⁴⁾ 7i^r 7ya¹³ kyee^{31(r)} s7e^r nskwa² wa7³⁴ni⁰⁽³⁴⁾.
 go.out.COMPL-lie.PROG to near stone place lie.HAB now
 ‘It came out lying on a stone, where it is to this day.’
 Chatino (McIntosh 2011: 117)

Since the use of *skaʳ* seems restricted to referents of typical presentational contexts and other non-identifiable referents that are prominent within a given discourse sequence, I treat *skaʳ* in Chatino as a presentational article. Being a presentational article, *skaʳ* is not restricted to specific indefinite referents. The example in (452) shows that *skaʳ* also occurs in contexts that allow for a nonspecific interpretation of the referent. Another nonspecific context in (453) on the other hand shows that a nonspecific referent can also be expressed by a bare noun.

- (452) ja^r n7i^r ska^r nte²⁽⁺⁰⁾ re³⁴.
 NEG live.HAB ART:PRES people DEM
 ‘A person doesn’t live here.’
 Chatino (McIntosh 2011: 104)

- (453) s7i²⁽⁺⁰⁾ knya7³¹ ka^r nu^r mn7a²¹.
 NEG deer be.PROG REL see.COMPL.2SG
 ‘You did not see a deer.’
 Chatino (McIntosh 2011: 104)

Example (454) illustrates another use of *skaʳ*: it occurs with a nominal predicate, which means that the specific article may also be used in nonreferential contexts.

- (454) *jychɛr* *7ya*³⁴-*j7o*^{31(t)} *ka*^r *ra*^r *ska*^r *jychɛr* *tlyu*¹³.
 village mountain-holy be.PROG OBJ [ART:PRES village big]
 ‘Santa Lucía is a big town.’
 Chatino (McIntosh 2011: 99)

Again, such uses emphasize the importance of separating the two extension processes in the development of articles, i.e. the discourse-pragmatic extension to less prominent referents and the semantic extension to nonspecific referents or, in this case, non-referential expressions.

6.2.4 Kashibo-Kakataibo

Kashibo-Kakataibo is a Panoan language spoken in the Huánuco and Ucayali departments of Peru. Zariquiey Biondi (2011: xiv) estimates that the language has 3000-5000 speakers. Kashibo-Kakataibo only has a presentational article and no other articles in the definite domain.

What I call presentational article here is the marker *achushi*, homophonous with the numeral ‘one’. Zariquiey Biondi (2011: 302-303) notes that *achushi* can be interpreted as the numeral ‘one’ or as an indefinite marker depending on its position relative to the nominal head noun. If it precedes the noun, it marks the referent of the head noun as indefinite, and if it follows the noun, it is used as the numeral ‘one’.⁶ In the latter position, it is also often used with the meaning of ‘another’, contrasting it with another referent of the same kind. Examples of those two uses are shown in (455) and (456). In both examples, the referent of *uni* ‘person’ is marked by *achushi* preceding the noun, introducing the referent to the discourse. The referent of *maxax* ‘stone’ in (456) is contrasted with another previously mentioned stone; in this case, *achushi* marks this contrast and follows the noun.

- (455) *a=x* *kaisa* *'i-akë-x-a* [*achushi uni*] *a-n*
 3SG=S NARR.REPORT.3 be-REMPST-S-NON.PROX ART:PRES person 3SG=ERG
 ‘axan-kë.
 fish.using.poison-NOM
 ‘It is said that he was a man who used to fish using poison.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 744)

⁶Zariquiey Biondi (2011) nevertheless often translates *achushi* as ‘one’ in the examples. I follow the original translation, even though it may not necessarily reflect the referential function of *achushi*.

- (456) a=mi nish-kin kaisa [achushi uni=n] maxax maxax=nu ain
 3SG=LOC envy-SS NARR.REPORT.3 ART:PRES person=ERG stone stone=LOC GEN.3
 bëru nan-kë anu bëru nan-kë kaisa kwan-xun [maxax achushi=n]
 eye.ABS put-NOM there eye.ABS put-NOM NARR.REPORT.3 go-SS stone one=INS
 chaka-akë-x-ín.
 beat-REMPST-3-PROX
 ‘It is said that envying him one man, going to the place the man had put his eye on a stone,
 beat the eye with (another) stone.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 735)

More evidence for prenominal *achushi* no longer being a numeral but a referential marker comes from the (in)compatibility of *achushi* with demonstratives. If used postnominally as a numeral, it can co-occur with an adnominal demonstrative, as in (457a). However, this is not the case with prenominal *achushi*, as is shown in (457b).

- (457) a. ënë uni achushi
 DEM man one
 ‘this one man’
 b. *ënë achushi uni
 DEM ART:PRES man
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 308)

Note that the incompatibility of the demonstrative in (457b) and *achushi* is not due to a restriction against two prenominal modifiers; example (458) shows such a configuration with a demonstrative and an adjective.

- (458) ënë upí xanu
 DEM beautiful woman
 ‘this beautiful woman’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 308)

Other examples of *achushi*, introducing discourse-prominent and indefinite referents are given in (459) and (460), and (461). Example (459) shows again that the presentational article is used to introduce the protagonist at the beginning of a story. In (460), the referent of *xëxat* ‘river’, marked by *achushi* as well, is not a protagonist but nevertheless a prominent referent in the discourse, as it sets the scene for the following sequence of the story.

- (459) [achushi uni] kaisa u-akë-x-a atu=nu
 ART:PRES person.ABS NARR.REPORT.3PL come-REMPST-3-NON.PROX 3PL=LOC
 bëba-akë-x-ín.
 arrive-REMPST-S-PROX
 ‘It is said that one man arrived to where they used to live.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 756)

- (460) kwan-xun kaisa is-akë-x-ín [achushi xëxat=nu] nukut-xun
 go-SS NARR.REPORT.3 see-REMPST-3-PROX ART:PRES small.river=LOC arrive-SS
 ain amo 'i-kë bëru tsëka-xun maxax=nu nan-ia is-akë-x-ín.
 GEN.3 at.the.other.side be-NOM eye.ABS take.off-SS stone=LOC put-SS see-REMPST-3-PROX
 'Going, they saw (him) arriving at one small river, taking out one of his eyes and putting
 it on a stone.'
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 748)

In (461), we see *achushi* marking a referent as indefinite which will become the new center of attention. The center of attention of the preceding discourse segment is the *gringo*, who died. In the following discourse segment, the new center of attention is a banana tree (*moquicho*) growing from the heart of the *gringo*. Therefore, we see in (461c) that the referent of *shinkun* 'tree of *moquicho*' is marked by *achushi* to signal that it is discourse-prominent but not (yet) identifiable to the hearer.

- (461) a. a uni a gringo a bëruan-këx 'i-këbë=bi
 that person.ABS that white.person O:3SG take.care-SS be-DS=same
 'Even though they had taken care of that *gringo*,'
- b. dentro.de.un.mes kaisa medio.mes kaisa atian ënu a=x kan
 after.one.month NARR.REPORT.3 half.a.month NARR.REPORT.3 then here that=S PART
 ain nuitu puntë ënu-ax
 3SG.GEN heart straight here-s
 'after one month or half a month, here from his heart, here,'
- c. kaisa [achushi shinkun shinkun]
 NARR.REPORT.3 ART:PRES banana.species.ABS banana.species.ABS
 uni-akë-x-ín.
 spring.up-REMPST-3-PROX
 'one tree of *moquicho* [a banana species] sprang up.'
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 761)

Example (462) and (463) are the two utterances that follow (461) in the story. In (462), we see that the banana tree is indeed the discourse topic, and (463) shows how the next referent or the new discourse topic is introduced. The speaker tells the hearer about a long bunch of bananas that grew from the banana tree. Again, the discourse-prominent referent of *racimo* 'bunch' is introduced by the presentational article *achushi*.

- (462) "shinkun ka uni-i-a kantsin ka
 banana.species.ABS NARR.3 spring.up-IPFV-NON.PROX banana.species.ABS NARR.3
 uni-i-a" ki-xun kaisa is-akë-x-ín.
 spring.up-IPFV-NON.PROX say-SS NARR.REPORT.3 see-REMPST-3-PROX
 'The people saw it, saying "a *moquicho* tree has grown".'
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 762)

- (463) a. is-kin bėruan-kė bėruan-kė
 see-ss take.care-NOM take.care-NOM
 ‘From the corpse that they had looked after for a long time,’
- b. kwan-ax kwan-ru-tankėx manan=mi kwan-ru-tankėx [**achushi** racimo]
 go-ss go-up-ss upside.of=LOC go-up-ss ART:PRES bunch
 chaxkėė=ira como.tres.metros usa-i kaisa a shinkun
 long-INTERR like.three.meters like.that-ss NARR-REPORT.3 that banana.species.ABS
 bakėn-akė-x-ın bimi-akė-x-ın.
 give.birth-REMPST-3-PROX get.fruit-REMPST-3-PROX
 ‘a tree started to grow very high and it got a very long bunch of bananas, like three meters long.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 762)

Two other example of *achushi* marking a discourse-prominent referent as indefinite are given in (464) and (465). These two contexts show that *achushi* is also used outside of story telling, where the discourse referents is simply relevant in the current discourse situation. In (464), the referent of *matá* ‘hill’ is specific; in (465), on the other hand, we see that *achushi* can also occur with nonspecific referents.

- (464) a=x ka ’ikėn [**achusi** matá] ka is!
 that=s 3:NARR be.3 ART:PRES hill.ABS NARR look
 ‘There is a **hill**, look!’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 250)

- (465) ’ė=x Lima=nu kwan-xun kana [**achushi** casaca] bits-kė ’i-tsin-a-n.
 1SG=s Lima=LOC go-ss 1SG:NARR ART:PRES jacket buy-NMLZ be-COND-PFV-1
 ‘If I had gone to Lima, I would have bought a **jacket**.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 436)

Examples (466) and (467) show that the presentational article *achushi* is not always used to mark specific referents. Both referents of *bai* ‘path’ and *kapė* ‘caiman’ are expressed without the article.

- (466) [cha **bai**] ka ’a-akė-x-a desde Tingo Maria anuxun [...] hasta
 big **path**.ABS NARR.3 do-REMPST-3-NON.PROX from Tingo Maria there [...] until
 Pucallpa=nu.
 Pucallpa=LOC
 ‘They made a big road from Tingo Maria up to Pucallpa.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 320)

- (467) a munu ka nit-i=ma nukën nanët-baë=n kamënë [kapé kamó]
 that.O slowly NARR.3 walk-SS=NEG 1PL.GEN brother-COLL=ERG NARR.3.MIR caiman big
 ‘a-xun ain të-xaká rakan-bian-on-x-a.
 do-SS 3SG.GEN neck-hide.ABS lean-going-PST.day.before-3-NON.PROX
 ‘Look! Without walking slowly, our brothers, killing a big caiman, have left its neck hide.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 438)

In (468) and (469), the referents of *bakan* ‘wasp’ and *maxax* ‘stone’ are also specific but clearly not relevant to the story; they are expressed by bare nouns as well.

- (468) [...] kwënkën-but-kian-i kwan-akë-x-a uni ñusi a=x bakan bina
 [...] shout-down-going-SS go-REMPST-3-NON.PROX person old 3SG=S wasp
 këñu-këx.
 finish-DS
 ‘[It is said that, when the pomegranate sank down, going together, when the waves carried them far,] the old man went shouting, because a wasp stung him completely.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 413)

- (469) kwan-kin kaisa kapě kapě rët-xun ain të-xakat maxax=nu
 go-SS NARR.REPORT.3 caiman caiman.ABS kill-SS 3SG.GEN neck-skin.ABS stone=LOC
 rakan-bian-akë-x-ín.
 lay.down-going-REMPST-3-PROX
 ‘It is said that, going, killing several caimans, [they] laid down its neck skin on a stone and thus went.’
 Kashibo-Kakataibo (Zariquiey Biondi 2011: 412)

6.2.5 Presentational articles and the development of NUM > ART:INDEF

The examples from Lango, Bilua, Chatino, and Kashibo-Kakataibo showed that we need to distinguish presentational articles as indefinite articles which are restricted to discourse-prominent referents. Depending on additional language-specific factors, this may result in a strong association with certain syntactic functions and/or positions in the clause. The presentational article in Lango was shown to be restricted in this way to occur with subjects only. The data from the other languages presented in this section showed that presentational articles are not necessarily syntactically restricted in that way.

The general restriction to mark only discourse-prominent referents may suggest that presentational articles are used rather infrequently, e.g. when compared with the distribution of indefinite articles in other languages. In this vein, Noonan (1992: 162) explicitly notes that most nouns in Lango are not marked as indefinite by *-mórô*. It may therefore be tempting to analyse presentational articles as optional markers and not to treat them as articles. One important point of

this section was to show that frequency of occurrence is independent from optionality or systematicity. Their likely less frequent use compared to indefinite articles in other languages does not make presentational articles optional; they are nevertheless used systematically to mark discourse-prominent referents as indefinite or non-identifiable (by the hearer). Of course, a corpus study allowing for a quantitative analysis of the distribution of the marker would be necessary to properly analyse to what extent the marker is used systematically with discourse-prominent referents. Or, grammaticality judgments by native speakers could be used to evaluate how felicitous the absence of the marker with indefinite but discourse-prominent referents is. Then, if we find that the marker generally occurs with discourse-prominent indefinite referents, and if its absence with prominent indefinite referents is judged as infelicitous by native speakers, the marker is a presentational article rather than a numeral (or another type of marker) with extended referential functions. Thus, presentational articles do not necessarily have to be frequent, and we should be careful not to assume a default distribution similar to the one of the indefinite article in e.g. English.

The other important point of this section was to show that presentational articles are not restricted to occur with specific referents, but that they can be used to mark nonspecific referents as well. This observation suggests that the pragmatic restriction to the occurrence with discourse-prominent referents is independent of the semantic restriction to specific contexts. However, the grammaticalization path from the numeral ‘one’ to the indefinite article appears to imply the contrary. Figure 6.1 shows the development proposed by Heine et al. (1995) introduced in Section 4.3.

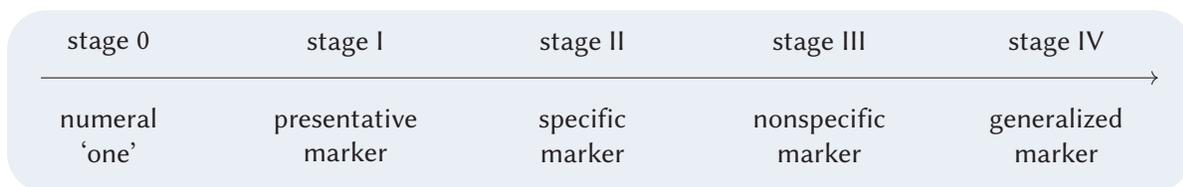


Figure 6.1: Grammaticalization path from the numeral ‘one’ to a nominal marker

Indefinite articles correspond to stage III in Figure 6.1, since they systematically encode specific and nonspecific referents. According to this path, the first necessary functional extension of the numeral is the introduction of discourse-prominent referents to the discourse, i.e. its “presentative function” in stage I.⁷

In order to develop further into an article, it is assumed that the use of the stage I marker then becomes more frequent and conventionalized so that it extends to specific referents which are

⁷However, as was shown in Section 4.3, evidence from languages with no systematic indefinite marker can use the numeral ‘one’ in certain contexts. This suggests that, in fact, the numeral ‘one’ may generally be able to take on the function of a stage I “presentative marker” if no other indefinite marker is available in the language.

not discourse-prominent referents. Section 6.3 shows examples of such stage II markers, called exclusive-specific articles in the present study. The development shown in Figure 6.1 seems to imply that an extension from specific to nonspecific referents only happens after the marker has developed into a stage II article. In other words, the pragmatic constraints of the stage I marker have to disappear first, and only then is the use of the marker semantically extended to nonspecific referents in stage III. However, the markers described as presentational articles in this section showed that pragmatic constraints can still apply to markers encoding both specific and nonspecific referents. Because most discourse-prominent indefinite referents are specific and not nonspecific, the use of the numeral ‘one’ or a presentational article with a specific referent is much more likely than their use with a nonspecific referent. This is especially true for contexts in which new referents are introduced. As a consequence, early specific, presentational, or indefinite articles do probably first occur with specific referents rather than nonspecific ones. However, we cannot maintain that a marker would have to “fully extend” its use with specific referents first before being used in nonspecific contexts.

Another related point, which is not necessarily implied in the path in Figure 6.1, though, is that presentational articles should not be analysed as a merely intermediate stage on the development towards a “fully-fledged” indefinite article. We do not know that articles restricted to discourse-prominent referents would necessarily be “unstable” across time; it is equally possible that presentational articles never extend their use to less prominent discourse referents.

6.3 Exclusive-specific articles

This section deals with exclusive-specific articles. Exclusive-specific articles encode a specific referent and cannot be used to mark definite or nonspecific referents. Many of the examples in this section will consist of contexts that are set in the past, including a referent that is marked as non-identifiable by the speaker, as in *I was attacked by a penguin yesterday*. Following Giannakidou (1997, 2017), I take contexts that refer to single past events as episodic contexts and the non-identifiable referent that is involved as specific.

6.3.1 Biak

Biak is an Eastern Malayo-Polynesian language spoken in Indonesia on the Bird’s Head Peninsula of Papua and on the islands Biak, Supiori, Numfor, and other smaller surrounding islands (van den Heuvel 2006: 1). According to the Endangered Languages Project, Biak has approximately 70,000 speakers.⁸

⁸Cf. <http://endangeredlanguages.com/lang/1363>.

Besides the exclusive-specific article, Biak has a definite article and a nonspecific article. This section will present the distribution of the exclusive-specific article in relation to the other two articles. The article has different exponents depending on the number value of the noun it occurs with; in the plural, the animacy of the referent is marked on the article as well. Table 6.1 gives an overview of the different realizations of the exclusive-specific article (van den Heuvel 2006: 66, 72). Van den Heuvel (2006: 204) explains that the forms with final *-ya* can be used in all contexts, while the forms ending in *-i* are restricted to prepausal positions. In the text, I will refer to the exclusive-specific article as *=ya* for the sake of simplicity.

Table 6.1: Exponents of the Biak exclusive-specific article

singular	<i>=ya/=i</i>
dual	<i>=suya/=sui</i>
trial	<i>=skoya/=skoi</i>
plural animate	<i>=sya/=si</i>
plural inanimate	<i>=na</i>

Examples (470) and (471) illustrate the use of the exclusive-specific article *=ya* to introduce new discourse referents. In (471a), the referent of *mankroder* ‘frog’ is mentioned for the first time in the discourse and marked as specific by *=ya*. When it is mentioned again in (471b), it is marked by the definite article *anya*.

- (470) Ras oso harimau=*ya* d-ors ro apyadwar=*i*.
 day some tiger=**ART:EXSPEC** 3SG-stand LOC cage=**ART:EXSPEC**
 ‘One day, a tiger stood in a cage.’
 Biak (van den Heuvel 2006: 89)

- (471) a. Indya fyanu rofan *anya*, i-fnovku mankroder=*i*.
 so feed.3SG dog **ART:DEF** 3SG-with frog=**ART:EXSPEC**
 ‘So he took care of the dog, together with a **frog**.’
- b. [mankroder *anya*] ma romá-mkun *anya* d-úf i.
 frog **ART:DEF** TOP child-little **ART:DEF** 3SG-pick.up 3SG
 ‘As for the frog, the little child picked it up.’
 Biak (van den Heuvel 2006: 204-205)

Examples (472), (473), and (474) show three more instances of *=ya* with a specific referent. Marking the referents of *wai mkun* ‘little canoe’, *for veba* ‘big fire’, and *nyan* ‘road’ in these examples as specific, the speaker also indicates that the exact identity of the referents is not relevant.

- (472) Ya-vors [wai mkun=*ya*] ro di-ne.
 1SG-row canoe little=**ART:EXSPEC** LOC place-here
 ‘I rowed a little canoe here.’
 Biak (van den Heuvel 2006: 108)
- (473) I-frúr [for ve-ba=*ya*].
 3SG-make fire REL-big=**ART:EXSPEC**
 ‘He made a big fire.’
 Biak (van den Heuvel 2006: 68)
- (474) Mura ido evekasip mos ani-pur-wa d-ák-smai
 afterwards THEME REL.VBLZ.scabies dirty ART:DEF-back-over.there 3SG-also-acquire
 nyan=*ya* fa d-óve ah, bo imbo.
 road=**ART:EXSPEC** CONS 3SG-say ah EXCL indeed
 ‘Afterwards the dirty one having scabies in the back also found a way out and thought ‘ah, indeed.’’
 Biak (van den Heuvel 2006: 432)

We see another example of the exclusive-specific article =*ya* in (475). Besides introducing =*ya* as a specific referent in the discourse in (475a), the article also signals that its identity is irrelevant for the following discourse sequence. The center of attention in both (475a) and (475b) is a heron, while the specific referent of *karui orovaido wonon* ‘stone or coral stone’, marked by =*ya*, is only briefly mentioned and not taken up in the following discourse segment. In (475a), we also see an anaphoric referent marked by the definite article *anya*. Example (475b) features the two situationally unique referents *wammurm* ‘wind from the East’ and *swan* ‘sea’. All three definite referents are marked by the definite article.

- (475) a. Imbarya man-sórom **anya** i-rov ma, irwai rar-vav vo kyain ro
 therefore bird-heron **ART:DEF** 3SG-fly to.here 3SG-land thither-down SIM 3SG.sit LOC
 [karui orovaido wonon=*ya*] bo-ri.
 stone or coral.stone=**ART:EXSPEC** upside=POSS:SG
 ‘So the heron flew close, he landed at sat down on a stone or coral stone.’
- b. Kyain vo, [wammurm **anya**] i-wuf pararei, [swan **anya**]
 3SG.sit SIM wind.east **ART:DEF** 3SG-blow healthy.RED sea **ART:DEF**
 bo-ri i=no i-vrin, mura ido myander vo d-óve ...
 upside-POSS.SG 3SG=also 3SG=quiet afterwards THEME 3SG.long.for SIM 3SG-say ...
 ‘He sat while the wind from the East was blowing gently, the sea surface, it was also quiet, so he said longfollly: ...’
 (van den Heuvel 2006: 217-218)

The exclusive-specific article can also occur with possessives, shown in (476). Here, =*ya* occurs with *srar vyed* ‘his sister’ and marks the referent as specific.

- (476) Srar vye=d-ya fyarvuk.
 cross.sibling POSS:3SG=3SG-ART:EXSPEC marry.3SG
 ‘His sister got married.’
 Biak (van den Heuvel 2006: 85)

The counterpart of the exclusive-specific article in the indefinite domain is the nonspecific article =o. The following examples show contexts in which the referent is necessarily interpreted as nonspecific. We see that the nonspecific article is required, and that the exclusive-specific article =ya is not used. In (477) and (478), the nonspecific referents marked by =o are part of a negated existential construction.

- (477) sansun vye=na na-is-ya voi, d-ák-nayu
 clothes POSS:3SG=ART:SPEC.3PL.INAN 3PL.INAN-PRED-that but 3SG-also-have
 sarak=o va.
 bracelet=ART:NSPEC not
 ‘His clothes were there, but he did not (also) have a bracelet.’
 Biak (van den Heuvel 2006: 400)

- (478) S-na mov=o ro di-wa fa s-om va.
 3PL.AN-have place=ART:NSPEC LOC place-over.there CONS 3PL.AN-clear.away not
 ‘They do not have a place there to make a garden (lit. to clear up a place).’
 Biak (van den Heuvel 2006: 129)

Example (479) also shows a context involving negation. Van den Heuvel (2006: 210) notes that “[...] the narrator expresses that there was not any man (*snon*) with whom the girl has slept. It is the non-existence of this man which accounts for the use of =o [...]”. Therefore, we can conclude that =o cannot scope above the negation operator.

- (479) I-snefr snon=o va vo, sne-ri i-ba ri.
 3SG-sleep.with male=ART:NSPEC not SIM belly-POSS 3SG-big ANAPH
 ‘Although she had not slept with a man, she was pregnant.’
 Biak (van den Heuvel 2006: 210)

Example (480) shows a non-episodic context. The referent of *vyemnór* ‘snotty nose’ is therefore nonspecific and marked as such by the use of the nonspecific article. The same article can be seen in (481), where the existence of the nonspecific referent of *romawa Papua oser* ‘a Papuan’ is within the scope of the question.

- (480) Oso vye-mnór=o i-pok vyunk awer mnór
 some 3SG.VBLZ-mucus=ART:NSPEC 3SG-be.able.to wipe.off.3SG PROHIB mucus
 vye=na.
 POSS:3SG=ART:EXSPEC.PL.INAN
 ‘If one has a snotty nose, he is not allowed to wipe off his mucus.’
 Biak (van den Heuvel 2006: 132)

- (481) Manseren, na da ya-roro nanki ya-mar ido ya-srow [romawa Papua
 Lord then perhaps 1SG-LOC heaven 1SG-die THEME 1SG-meet child Papua
 oser=**o**] ke?
 one=**ART:NSPEC** DOUBT
 ‘Lord, perhaps when I come into heaven, shall I maybe meet a Papuan?’
 Biak (van den Heuvel 2006: 210)

The contrast between the exclusive-specific article =*ya* and the nonspecific article =*o* shows an interesting effect resulting from the (lack of the) existence presupposition. This is shown in the two minimal pair contexts in (482) and (483). In (482a) and (483a), the specific article is used with the referents in question in the position of the direct object of the verb *fúr* ‘make’. Due to the existence presupposition that the exclusive-specific article comes with, the event is interpreted as a past event whose result is an existing ‘ukelele’ or ‘tea’. By contrast, in (482) and (483) the referents are marked as nonspecific through the use of the nonspecific article =(y)*o*. Because =(y)*o* does not have an existence presupposition, the events expressed by the two utterances receive a future interpretation.

- (482) a. I-fúr yuk=**ya** fa y-ún i ve Waranda.
 3SG-make ukelele=**ART:EXSPEC** CONS 1SG-take 3SG to the.Netherlands
 ‘He has made a ukelele so that I can take it to the Netherlands.’
 b. I-fúr yuk=**o** fa y-ún i ve Waranda.
 3SG-make ukelele=**ART:NSPEC** CONS 1SG-take 3SG to the.Netherlands
 ‘He is making / will make a ukelele so that I can take it to the Netherlands.’
 Biak (van den Heuvel 2006: 71)

- (483) a. Ya-fúr te=**ya** fa kuy-inm.
 1SG-make tea=**ART:EXSPEC** CONS 1DU.IN-drink
 ‘I (had) made tea for us so that we two can drink.’
 b. Ya-fúr te=**yo** fa kuy-inm.
 1SG-make tea=**ART:NSPEC** CONS 1DU.IN-drink
 ‘I am making tea for us so that we two can drink.’
 Biak (van den Heuvel 2006: 206)

To conclude, this section showed that Biak systematically uses =*ya* to encode specific referents. In definite contexts, we saw that the definite article *anya* is used instead. In nonspecific contexts Biak uses the nonspecific article =*o*. Thus, Biak =*ya* has to be analysed as an exclusive-specific article.

6.3.2 Akan

Besides the anaphoric article presented in Section 5.2.3, Akan (Kwa, Ghana) also has an exclusive-specific article, *bi*, which was discussed as a marker of specificity in the previous literature, most notably in Arkoh (2011). This section shows how *bi* is systematically used in specific indefinite contexts but does not occur with definite or nonspecific referents. In (484), we see *bi* in the typical specific contexts, introducing a new discourse referent:

- (484) Da bi [ɔbea bi] ne ne ba ɔsoɔdenfo bi tena-a
 day ART:EXSPEC woman ART:EXSPEC CONJ POSS child stubborn ART:EXSPEC stay-COMPL
 ase.
 under
 ‘Once upon a time, there was a woman and her stubborn child.’
 Akan (Amfo 2010: 1791)

The following examples (485), (486), and (487) show that the exclusive-specific article is used in other types of contexts involving specific referents as well. In these examples, in contrast to (484), the specific referent has less discourse prominence. The article *bi* occurs together with the expressions *akyɛde* ‘gift’, *ahwahwadeɛ* ‘miracle’, and *mayɔnkofɔɔ* ‘my friends’, which are all linked to a particular referent that is not identifiable by the hearer. In (485), the referent becomes identifiable by the hearer at a later point in time; the context for (486) excludes the hearer from knowing about the ‘gift’, and the identifiability of the specific referent in (487) is simply not relevant in the given situation.

- (485) Nanso [ahwahwadeɛ *(bi)] si-ɛ ɔbaabunu no fa-a afuro.
 but miracle ART:EXSPEC happen-PST virgin ART:DEF take-PST stomach
 ‘But a miracle occurred. The virgin conceived.’
 Akan (primary data)
- (486) Me-wɔ [akyɛde *(bi)] ma wo.
 1SG-have gift ART:EXSPEC give 2SG
 ‘I have a gift for you.’
 Akan (primary data)
- (487) [M-ayɔnkofɔɔ *(bi)] ba-a me-hɔ ɛnora anadwo.
 POSS:1SG-friends ART:EXSPEC come-PST POSS:1SG-place yesterday night
 ‘Some friends came over to my place last night.’
 Akan (primary data)

Examples (488) and (489) show that *bi* cannot be used with nonspecific referents. In (488), the referent of *yere* ‘wife’ is nonspecific because it is within the scope of the irrealis reading of the predicate; (489) shows a nonspecific referent in the context of a question.

(488) Me-nya-a anka me-wɔ [yere (*bi)].
 1SG-get-PST before 1SG-have wife ART:EXSPEC
 ‘I wish I had a (any) wife.’
 Akan (primary data)

(489) Wo-bɛ-nya [pen (*bi)] a-ma me anaa?
 2SG-FUT-get pen ART:EXSPEC IMPERS-give 1SG Q
 ‘Can you give me a (any) pen?’
 Akan (primary data)

In order for *bi* to be an exclusive-specific article, it needs to be incompatible with definite referents. That Akan uses the separate anaphoric article *nó* in definite contexts was extensively discussed in Section section 5.2.3. Therefore, we can conclude that Akan *bi* is indeed an exclusive-specific article.

6.3.3 Palula

Palula (Dardic, Pakistan) is another language with an exclusive-specific article. Examples (490) to (493) show that the article *áa/ák*⁹ marks the referent as specific, i.e. as particular but not identifiable by the hearer.

(490) bhuná [áa gíri] heensíl-i
 below ART:EXSPEC rock stay.PFV-F
 ‘Down below there was a big rock.’
 Palula (Liljegren 2016: 309)

(491) miír thaní [áak míiš] heensíl-u de.
 Mir QUOT ART:EXSPEC man stay.PFV-M.SG PST
 ‘There was a man called Mir.’
 Palula (Liljegren 2016: 309)

While examples (490) and (491) illustrated the use of the exclusive-specific article in the typical specific contexts of the introduction of a new, prominent discourse referent, (492) and (493) below show that it is also used with less prominent specific discourse referents.

(492) eesé zangal-í [áa baṭ-á] ḡhulí hari so kuṇaák
 REM forest-OBL ART:EXSPEC stone-OBL on take.away.CV DEF.NOM.M.SG child.NOM.M.SG
 bheešóol-u.
 seat.PFV-M.SG
 ‘In that forest he took the child to a stone and seated him.’
 Palula (Liljegren 2016: 283)

⁹The different forms are used in different varieties of Palula (Liljegren 2016: 140).

- (493) [ak táaper-e] ték-a
 ART:EXSPEC hill-GEN top-OBL
 ‘on the top of a hill’
 Palula (Liljegren 2016: 281)

The article *áa/ák* does not occur with nonspecific referents. Examples (494) to (496) below show three contexts with nonspecific referents which are all expressed as bare nouns, namely *fláit* ‘flight’, *wása* ‘strength’, and *muloó* ‘mullah’, respectively.

- (494) čúur reet-í jheez-í fláit na bhíl-i hín-i aáj bi kansál
 four night-PL airplane-GEN flight NEG become.PFV-F be.PRS-F today also cancelled
 bhíl-i.
 become.PFV-F
 ‘There have been no flights for four days, and also today it was cancelled.’
 Palula (Liljegren 2016: 243)

- (495) tasí ba ga wása na heensíl-u.
 3SG.GEN TOP any strength NEG stay.PFV-M.SG
 ‘And he had no strength at all.’
 Palula (Liljegren 2016: 310)

- (496) muloó díi yeér kráam na bháan-u.
 mullah from without work NEG become.PRS-M.SG
 ‘Without a mullah the work is not being done.’
 Palula (Liljegren 2016: 332)

Neither can the exclusive-specific article *aa/ak* mark definite referents. Example (497) below shows an anaphoric referent which is expressed by a noun together with the anaphoric article *se*.

- (497) théeba [se thaataak-a] bi tas sangí khainí široó thíl-u.
 then ART:ANA monster-OBL also 3SG.ACC with eat start do.PFV-M.SG
 ‘(When the monster came inside, the man was eating.) And the monster started eating with him.’
 Palula (Liljegren 2016: 141)

6.4 Nonspecific articles

This section presents three examples of nonspecific articles. Nonspecific articles are articles that only occur with nonspecific but not with definite or specific indefinite referents. This article type is very rare crosslinguistically; my sample contains only 5 languages with nonspecific articles.

Section 6.4.1 presents the nonspecific article in Ayoreo, whose article system also includes an inclusive-specific article. Section 6.4.2 deals with the nonspecific article in Lakota. Lakota has a more complex article system consisting of a definite article (cf. Section 5.3.3), an exclusive-specific article, and the nonspecific article discussed here. The third example the nonspecific article in Tongan, discussed in Section section 6.4.3. The Tongan article system includes another inclusive-specific article and an anaphoric article, both of which will be presented in more detail in Section 7.1.3. Finally, Section 6.4.4 discusses the development of nonspecific markers in a number of languages from North America, arguing for their development from verbal irrealis markers.

6.4.1 Ayoreo

Ayoreo is a Zamucoan language spoken in the Chaco Boreal between Bolivia and Paraguay. The language has approximately 3000 speakers.¹⁰ The article system in Ayoreo consists of an inclusive-specific and a nonspecific article. The former expresses definite and specific referents, while the latter is used to mark nonspecific referents.

What I treat as a nonspecific article here is called the “indeterminate form” in Bertinetto (2009) and Ciucci (2016) as opposed to the “base form” and the “full form”. The base form corresponds to the root of a noun in Ayoreo (Ciucci 2016: 444). Nouns are realized in their base form if used as predicates. When they are used as arguments and are referential, they need to occur in their “full form”, which corresponds to the base form with a suffix marked for gender (masculine vs. feminine) and number (singular vs. plural). However, this form does not occur with nonspecific referents; in those contexts, another so-called “indeterminate form” is used. As Bertinetto (2009: 23) notes: “The label ‘indeterminate’ appears to be, in this case, well-chosen, for this form always implies a non-specific referent.” Therefore, I treat the full form as a noun with an inclusive-specific article and the indeterminate form as a noun with a non-specific article. The forms of the both articles are shown in Table 6.2 following Ciucci (2016: 447,458,476-481). Due to different assimilation and harmony processes as well as irregular forms, the exponents given in Table 6.2 are only the most common forms and not a comprehensive list.

Table 6.2: Ayoreo article exponents

		inclusive-specific	nonspecific
FEM	SG	-a/-ia/-∅	-raque/-taque/-naque
	PL	-die/-i	-rigi/-tigi/-niŋi
MASC	SG	-i	-tique/-rique/-nique
	PL	-ode	-tigo/-rigo/-niŋo

¹⁰Cf. <http://endangeredlanguages.com/lang/685>.

Examples (498) and (499) contrast the use of the base form in a predicative context with the use of the inclusive-specific article with a definite referent. The base form of *carataque* ‘jaguar’ is used as a predicate in (498), while the inclusive-specific article is required when expressing a definite referent as in (499).

(498) ¡Carataque que, don Pedro a!
 jaguar.M.SG PST don Pedro MOD
 ‘It was a jaguar, don Pedro!’
 Ayoreo (Bertinetto 2009: 40)

(499) Ch-uninga mu carata-i t-ōraja guesi.
 3-be.surprised but jaguar-ART:INSPEC.M.SG 3-throw.into outside
 ‘He was surprised when the jaguar came out.’
 Ayoreo (Ciucci 2016: 483)

Example (500) shows a minimal pair of a specific and a nonspecific referent.¹¹ In (500a), we see a specific referent which is expressed by the inclusive-specific article *-i*. The nonspecific article *-tique*, on the other hand, is not felicitous in this context. In (500b), we see the opposite pattern with a nonspecific referent: here, the nonspecific article is required, and the inclusive-specific article can no longer be used.

(500) a. María pota nona *ata-tique / ata-i uté uje
 María wants accompanies rich-ART:NSPEC.M.SG / rich-ART:INSPEC.M.SG that COMP
 chise dirica.
 meets yesterday
 ‘María wants to marry that rich man whom she met yesterday.’

b. María pota nona ata-tique / *ata-i mu cama
 María wants accompanies rich-ART:NSPEC.M.SG / rich-ART:INSPEC.M.SG but not.yet
 chimó.
 sees
 ‘María wants to marry a rich man, but she has not yet met him.’
 Ayoreo (Bertinetto 2009: 46-47)

Nonspecific articles are commonly found in contexts involving negation, especially in existential constructions, as well as contexts with irrealis or future reference. Examples (501), (502), and (503) show the use of the Ayoreo nonspecific article with negated existential expressions.

(501) Mu que etotigue-rique cuse muñi Dupade chataja.
 but NEG strength-ART:NSPEC.M.SG there.is but God helps
 ‘He fees rather weak (lit.: has no strength) but God helps him.’
 Ayoreo (Bertinetto 2009: 47)

¹¹In Bertinetto (2009: 46), the nonspecific article is spelled as *-tic* instead of *-tique* in example (500a) (but not in (500b) nor in any other example he provides of this marker). I adapted the spelling to the other examples.

(502) Que i-boca-**raque** cuse enga ch-ijna d-ojo-die.
 NEG 3-gun-**NSPEC.F.SG** EXIST and 3SG-bring 3.REFL-arrow-ART:**INSPEC.F.PL**
 ‘He does not have his gun, but he brings his arrows.’
 (Ciucci 2016: 484)

(503) Mu que ore i-plata-**rigui** cuse.
 but NEG 3PL 3-money-ART:**NSPEC.M.SG** EXIST
 ‘But they have no money.’
 Ayoreo (Ciucci 2016: 356)

Another example of the nonspecific article is shown in (504). Here, the article occurs with a nonspecific referent within the scope of a modal marker. Similarly, the nonspecific article in (505) is used with the noun *cucha* ‘thing’ in a counterfactual conditional construction.

(504) ujetiga adute cucha ajmacaca-**rique** maringa je ca atodo
 COMP listen.2PL.NON.IND thing ill.fated-ART:**NSPEC.M.PL** although MOD NEG fear.IMP
 cucha ajmacaca-rode.
 thing ill.fated-ART:**INSPEC.M.PL**
 ‘Even though you might hear threats, do not be afraid of them.’
 Ayoreo (Bertinetto 2009: 52-53)

(505) Ujetiga Jate di-rase nga, chisi-rase yogui-ji cucha-**rique**.
 COMP Jate arrive-MOD COORD give-MOD 1PL-LOC thing-ART:**NSPEC.M.SG**
 ‘If Jate arrived, he would give us something.’
 Ayoreo (Bertinetto 2009: 49)

The nonspecific article can also be used to express uncertainty about the identity of the referent. In (506), the speaker uses the nonspecific article to signal that they were not certain what kind of animal the referent refers to.

(506) A ore ch-ajna ajarame-**tique**. Aramoro-**raque** a
 MOD 3PL 3-follow armadillo-ART:**NSPEC.SG.M** brown.brocket-ART:**NSPEC.SG.F** MOD
 deji.
 EXIST.3
 ‘They are following an armadillo, or perhaps a brocket.’
 Ayoreo (Ciucci 2016: 484)

Another type of contexts that the nonspecific article is used in are contexts in which the speaker signals that the identity of the referent is not important, similarly to the free-choice use of e.g. *any* in English. This is shown in (507).

(507) Jirãque ore chise burica-**rigo**.
 suddenly 3PL meet horse-ART:**NSPEC.PL**
 ‘Suddenly they met (some) horses.’
 Ayoreo (Bertinetto 2009: 46)

Examples (508) and (509) show that contexts with a specific referent call for the use of the inclusive-specific article, confirming the pattern from example 500 at the beginning of this section.

(508) ¡Arócojna-quedejna-i deji ne!
 alligator-different-ART:INSPEC.M.SG EXIST.3 there
 ‘There is an alligator right there!’
 Ayoreo (Ciucci 2016: 485)

(509) Enga oriechoquí ch-ayója tié enga basá-ja [pararingué
 CONJ thief.ART:INSPEC.M.SG 3-run.to river CONJ 3.fall-into ravine.M.SG
 cucháb-i].
 big-ART:INSPEC.M.SG
 ‘The thief ran to the river and fell into a deep ravine’
 Ayoreo (Ciucci 2016: 485-486)

To conclude, we saw that the so-called indeterminate form of Ayoreo nouns systematically occurred to mark referents as nonspecific, while it was shown to be absent with definite and specific referents. Therefore, it must be analysed as a nonspecific article.

6.4.2 Lakota

Another example of a nonspecific article is found in Lakota (Siouan), whose definite article was already discussed in Section 5.3.3. Besides the definite article, *kiŋ*, Lakota has an exclusive-specific article *wā* which occurs in complementary distribution with the nonspecific article *wāži* in indefinite contexts.

Van Valin (1977: 64) discusses a further marker *cha*, which he calls an indefinite article. He notes that *cha* is usually not regarded as an article but that it “has the same general syntactic properties of the other articles”. Based on this language-internal argument, he treats it on a par with the other articles in Lakota. However, *cha* was described to encode contrast similarly to a focus marker (Van Valin 1977: 65-67, Williamson 1984: 49-50), which is why I do not treat it as an article here. This Section focuses on the nonspecific article and its distribution in relation to the exclusive-specific article.

An example contrasting the use of the nonspecific article with the use of the exclusive-specific article is given in (510). In constructions with intensional predicates, only the use of the exclusive-specific article *wā* allows for a specific interpretation of the referent of *igmu* ‘cat’. This is shown in (510a). Using *wāži* as in (510b) signals that *igmu* ‘cat’ has a nonspecific referent.

(510) a. [igmu’ wā] wachi.
 cat ART:EXSPEC want.s:1SG.O:3SG
 ‘I want a (particular) cat.’

- b. [igmu' wāži] wachi.
 cat ART:NSPSEC want.s:1SG.O:3SG
 'I want a (any) cat.'
 Lakota (Van Valin 1977: 64)

Since this distribution of *wā* and *wāži* with respect to specific and nonspecific contexts was reported to be systematic (Rood & Taylor 1996, Van Valin 1977, Williamson 1984), I treat *wāži* as a nonspecific article.¹² The full paradigm of the nonspecific and specific article is given in Table 6.3 below (Van Valin 1977: 64, Williamson 1984: 48, and Rood & Taylor 1996).

Table 6.3: Lakota exclusive-specific and nonspecific article paradigm

	exclusive-specific	nonspecific
SG	wā	wāži
PL	eya'	etā

Examples (511) to (514) further illustrate the use of the two articles in Lakota. In (511), the referent of *makno'xlova* 'cave' is specific and therefore marked with the specific article. In (512) and (513), the context does not disambiguate between a specific or a nonspecific interpretation of the referents *thaspā* 'apple' and *šuka* 'dog'. In these examples, it is the nonspecific article that resolves this ambiguity and marks the two referents as nonspecific. Example (514) shows the plural form *etā* of the nonspecific article.

- (511) [makno'xloka wā] e'l eū'papi.
 cave ART:EXSPEC in laid.s:3PL.O:3SG
 'They laid him in a cave.'
 Lakota (Van Valin 1977: 66)
- (512) [thaspā wāži] wachi.
 apple ART:NSPEC want.s:1SG.O:3SG
 'I want an (any) apple.'
 Lakota (Williamson 1984: 48)
- (513) [šuka wāži] ophethu maši.
 dog ART:NSPEC buy ask.s:3SG.O:1SG
 'He asked me to buy a (any) dog.'
 Lakota (Williamson 1984: 49)

¹²Even though these two articles encode specific and nonspecific referents, respectively, they are treated as two indefinite articles in Van Valin (1977: 63), as two indefinite determiners in Williamson (1984: 48), and as two indefinite markers in Ingham (2003: 90).

- (514) [ch̥ etā] aku wa!
 stick ART:NSPEC.PL bring IMP
 ‘Bring back some (any kind of) sticks!’
 Lakota (Williamson 1984: 48)

In (515), the nonspecific referent of *thaspā* ‘apple’ is within the scope of negation. In this case, the nonspecific article bears an additional negation marker and occurs as *wāži-ni*.

- (515) [thaspā wāži-ni] tebwaye šni.
 apple ART:NSPEC-NEG eat.up.1SG NEG
 ‘I didn’t eat up an (any) apple.’
 Lakota (Williamson 1984: 49)

Rood & Taylor (1996: 36) treat nonspecific articles that combine with the negation marker as in (515) as a separate negative nonspecific article and list additional exponents for combinations with non-human and inanimate plural nouns. However, examples of such nonspecific articles in negation contexts are rare in the literature on Lakota; the only marker that is shown in such contexts being *wāžini* (Ingham 2003, Rood & Taylor 1996, Van Valin 1977, Williamson 1984). Regardless of this issue, I do not treat the form *wāži-ni* in negation contexts as a separate type of articles but as the combination of the nonspecific article *wāži* and an additional polarity marker *ni* because negation itself is not a separate referential function.

Another type of contexts in which the nonspecific article is found is shown in (516) and (517). Here, the events depicted are not episodic, which makes the subjects nonspecific and requires the use of the nonspecific article *wāži* (spelled *waŋji*).

- (516) Tohaŋl [ziŋtkala waŋji] oc’aŋna waŋcag itazipa ihaŋke k’iyela waškite s’a.
 when bird ART:NSPEC shoot.when at.once bow end near notch ITER
 ‘Whenever they shot a bird, they would cut a notch near the end of the bow.’
 Lakota (Ingham 2003: 95-96)

- (517) na hehaŋl [ko’skalaka waŋji] t’awic’ut’uŋ haŋtaŋhaś wiŋyaŋ ki
 and then youth ART:NSPEC marry if woman ART:DEF
 t’uŋkaŋku ki túŋweni lila kic’i akic’iyuta šni.
 father.in.law.her ART:DEF never very with look.at.RECIP NEG
 ‘And then, when a young man married, his wife never looked much at her father in law and rather they were in avoidance relationship with each other.’
 Lakota (Ingham 2003: 97-98)

To conclude, we saw that *wāži* systematically marks a referent as nonspecific and is at the same time not able to occur with specific (nor with definite) referents. This warrants its status as nonspecific article.

6.4.3 Tongan

Tongan is an Oceanic language spoken in Tonga. It is one of the national languages in Tonga and has approximately 187,000 speakers.¹³ Following the reference grammars of Tongan by Churchward (1953) and Morton (1962), the Tongan article or determiner system has been the focus of a number of previous studies (Abner & Burnett 2010, Ahn 2016, Anderson & Otsuka 2006, Hendrick 2005, MacDonald 2014).¹⁴ While the details differ from account to account, all studies distinguish three relevant markers. The marker *ha*, the marker *(h)e*, and *(h)e* together with a stress shift (or vowel lengthening) towards the right edge of the nominal complex.

This Section will focus on *ha* and argue that it is a nonspecific article. In doing so, I follow Anderson & Otsuka (2006), MacDonald (2014), Völkel (2010), who either call *ha* a nonspecific article or explicitly note that it does not simply mark indefiniteness in general but occurs in nonspecific or non-referential contexts.¹⁵ The marker *(h)e* and the stress shift will be discussed in Section 7.1.3, where I argue that *(h)e* can be analysed as a weak inclusive-specific article, and *(h)e* together with the stress shift as an anaphoric article.

The paradigm of the articles in Tongan is given in Table 6.4 below (Völkel 2010: 159). Both the inclusive-specific and the nonspecific articles have two exponents, one used in a neutral way, the other one to mark affection. I will only discuss the markers that belong to the neutral series.

Table 6.4: Tongan article paradigm

	neutral	affective
inclusive-specific	(h)e	si'i
nonspecific	ha	si'a

Examples (518) and (519) illustrate the use of *ha* in contexts with modal operators introduced by imperatives, where the object is necessarily interpreted as having a nonspecific referent. Example (519) also shows that the nonspecific article is compatible with plural nouns.

- (518) Ha'u mo [ha afo].
 come with ART:NSPEC fishing.line
 'Bring a (any) fishing line.'
 Tongan (Churchward 1985: 25)

¹³Cf. <https://www.ethnologue.com/language/ton>.

¹⁴See MacDonald (2014: 39-43) for a detailed overview of the different analyses of *(h)e* and *ha* as well as additional accounts not mentioned here.

¹⁵Ahn (2016), Churchward (1953), Hendrick (2005), Morton (1962), Otsuka (2000) treat *ha* as an indefinite article or determiner. Abner & Burnett (2010) propose an analysis of *ha* as a numeral.

(519) Fai [ha 'u tohi].
 do.IMPER ART:NSPEC PL letter
 'Write (some) letters.'
 Tongan (Churchward 1985: 24)

In questions, referents marked by *ha* are also interpreted as nonspecific. This is shown in (520), (521), and (522).

(520) 'Oku 'i ai [ha maa]?
 PRS in there ART:NSPEC bread
 'Is there some bread?'
 Tongan (Otsuka 2000: 50)

(521) 'Oku 'i ai [ha vai mafana], pe 'ikai?
 PRES DAT there ART:NSPEC water warm OR NEG
 'Is there any warm water here, or not?'
 Tongan (MacDonald 2014: 40)

(522) Na'e lea [ha talavou]?
 PST spoke ART:NSPEC young.man
 'Did a young man speak?'
 Tongan (Hendrick 2005: 910)

Furthermore, in (523) we see that *ha* cannot be used with specific referents. In order to mark a referent as specific, the inclusive-specific article *he* must be used instead of *ha*.¹⁶

(523) Na'e sio 'a Sione ki [he / *ha ta'ahine].
 PST see ABS Sione to ART:INSPEC / ART:NSPEC girl
 'Sione saw a girl.'
 Tongan (Otsuka 2000: 50)

More evidence for *ha* marking the referent as nonspecific comes from expressions with intensional operators. In (524), (525), and (526), the contexts allow for the referents in question to be either specific or nonspecific. In the presence of *ha*, they can only be interpreted as nonspecific.

(524) 'Oku ou fie ma'u [ha vai mafana].
 PRS I want receive ART:NSPEC water warm
 'I want to receive warm water.'
 Tongan (Churchward 1985: 24)

(525) 'Oku kumi 'a Siaoisi ki [ha fefine potu].
 PRES look ABS Siaoisi for ART:NSPEC woman clever
 'Siaoisi is looking for a [any] clever woman.'
 Tongan (Hendrick 2005: 914)

¹⁶This is probably not entirely so, cf. examples (537), (538), (539), and (540), at the end of this Section.

- (526) ‘Oku pau keu lau [ha tohi].
 PRS must SUB.1SG read ART:NSPEC book
 Book report: ‘I have to read a book.’
 ≠ Assigned book: There’s a book I have to read.
 Tongan (Abner & Burnett 2010: 13)

The article *ha* also signals that referents are nonspecific in conditional constructions, where the event expressed is non-episodic. This is shown in (527) to (529).

- (527) Ka mate [ha tangata], teu ma‘u ha kāsalo.
 if die ART:NSPEC man FUT.1SG receive ART:NSPEC castle
 ‘If any [*a certain] man dies, I will receive a castle.’
 Tongan (Abner & Burnett 2010: 12-13)

- (528) Kapau ‘e lahi ange ‘i he vaeua ‘a e ‘u sea ‘oku ma‘u ‘e
 if FUT big more at ART:INSPEC half ABS ART:INSPEC PL seat PRES catch ERG
 [ha paati] ‘i he Falealea ...
 ART:NSPEC party in ART:INSPEC parliament ...
 ‘If a party captures more than half of the seats in parliament ...’
 Tongan (Hendrick 2005: 916)

- (529) Ko e hā te u fai kapau te u hiki ki [ha tu‘asila] fo‘ou ...?
 PRED ART:INSPEC what FUT I do if FUT I carry to ART:NSPEC address new ...
 ‘What do I do if I move to a new address ...?’
 Tongan (Hendrick 2005: 916)

Negation shows the same effect as the modal expressions, questions, and conditional constructions seen above. If used with *ha*, a direct object within the scope of negation can only be understood as having a nonspecific referent. This is shown in (530) and (531) and reflected by the use of *any* in the English translations. For (531), Abner & Burnett (2010: 2010) explicitly note that the referent of *tohi* ‘book’ cannot scope above the negation and be interpreted as a specific.

- (530) Na‘e ikai te u faka‘ita‘i [ha faiako].
 PST NEG PRED I irritate ART:NSPEC teacher
 ‘I didn’t irritate any teacher.’
 Tongan (Hendrick 2005: 916)

- (531) Na‘e ‘ikai keu lau [ha tohi].
 PST NEG SUB.1SG read ART:NSPEC book
 ‘I didn’t read any book.’
 ≠ ‘There is a book that I didn’t read.’
 Tongan (Abner & Burnett 2010: 13)

In addition to marking referents as nonspecific, *ha* can occur with nominal expressions in predicate positions. An example is given in (532). Here, *ha* occurs together with the nominal expression *faiako* as a nominal predicate marked by *ko*. The nonspecific article *ha* is also used with secondary predicates, as can be seen in (533). These are equally introduced by *ko* and require the use of *ha* with the nominal expression.

(532) Ko [ha faiako] ia.
 PRED ART:NSPEC teacher he
 ‘He is a teacher.’
 Tongan (Hendrick 2005: 911)

(533) ‘Oku tau lau ia ko [ha taki].
 PRES we regard him PRED ART:NSPEC leader
 ‘We regard him as a leader’
 Tongan (Hendrick 2005: 911)

Interestingly, MacDonald (2014: 52) notes that the use of *ha* is not felicitous in a context similar to (532). She provides the example given in (534) and mentions that “[m]y consultant’s explanation for rejecting *ha* was to say that there is only one cook in a restaurant” (MacDonald 2014: 52). This means that even though *e naohi kai* is translated by ‘a cook’ in (534b), the expression actually evokes a contextually unique referent which needs to be expressed by the inclusive-specific article. Thus, *e naohi kai* ‘a cook’ is very likely referential and not a nominal predicate in this case, and the entire construction is probably better characterised as an identificational construction. Given that, it is expected that the nonspecific article *ha* is not felicitous in (534b).

(534) a. Ko Sione ‘oku ne ngaane ‘i he fale kai ...
 PRED Sione PRS 3SG work DAT ART:INSPEC house food ...
 ‘Sione works in a restaurant ...’
 b. ... ko [e / *ha ngaohi kai].
 ... PRED ART:INSPEC / ART:NSPEC maker food
 ‘... he is a cook.’
 Tongan (MacDonald 2014: 52)

Furthermore, the distinction between specific and nonspecific markers in Tongan is also integrated into the possessive system. The exponents of adnominal possessives can be morphologically decomposed into three parts: the article, a possessive marker, and a person marker. This results in possessive markers that are sensitive to the specific-nonspecific distinction as well. There are two sets of adnominal possessive markers, one set (A-set) is used in syntactic S and A positions, the other set (O-set) is used in O positions (Völkel 2010: 160-162). An example of the use of possessive forms from these two sets is provided in (535). In both (535a) and (535b), possessive

markers from the O-set are used. In the question in (535a), we see the nonspecific possessive *hao*; its inclusive-specific counterpart would be *ho* (Völkel 2010: 161). The referent marked by a possessive in the answer in (535b), on the other hand is specific. It therefore occurs with the specific possessive *hoku*, whose nonspecific counterpart would be *haku* (Völkel 2010: 161).

- (535) a. ‘Oku ‘iai [hao tokoua]?
 PRES exist POSS:2SG.NSPEC sibling
 ‘Do you have a (any) sibling?’
- b. ‘Oku ‘iai [hoku tokoua].
 PRES exist POSS:1.INSPEC sibling
 ‘I have a sibling.’
 Tongan (Völkel 2010: 162)

Existential constructions are typical contexts for markers of specificity. In Tongan, one would expect the use of the inclusive-specific (*h)e* in such constructions. Interestingly, Hendrick (2005: 912) shows that *ha* can be used in existential constructions as well, which is rather incompatible with an analysis of *ha* as a nonspecific article, since nonspecific contexts do not include any existential presupposition. We see the example provided by Hendrick (2005) in (536), where *ha* is used to mark what appears to be the specific referent of *puaka* ‘pig’ in an existential construction. Referring to this example, MacDonald (2014: 45) notes the following: “Otsuka (p.c.) informs me that her and consultants’ intuitions are that (37a) [(536), LB] is more felicitous as a question, as to them *ha* is dispreferred in non-irrealis contexts.” This means that the use of *ha* in existential constructions with specific referents may not be judged as felicitous by all native speakers, and it may underlie dialectal (or another kind of) variation.

- (536) ‘Oku ‘i ai [ha puaka] ‘i Māketi.
 PRES DAT there ART:NSPEC pig DAT Market
 ‘There is a pig at the Market.’
 Tongan (MacDonald 2014: 45) adapted from Hendrick (2005: 912))

However, example (536) is not the only example in the literature where *ha* is used with specific referents, which makes it an issue that needs to be addressed. Other examples of *ha* with specific referents are given in (537) to (540). Example (537a) shows that *ha* can occur in the context of an intensional predicate without marking the referent as nonspecific. This becomes evident by the following utterance shown in (537b). A similar type of context is given in (538). Examples (539) and (540) do not make explicit that the referent marked by *ha* is specific with additional context, but both depict episodic past events which inherently only allow for a specific (or otherwise definite) interpretation of the referents in question.

- (537) a. ‘Oku kumi ‘a Siaosi ki [ha fefine poto].
 PRS look.for ABS Siaosi DAT ART:NSPEC woman wise
 ‘Siaosi is looking for a wise woman.’
 b. ‘Oku ‘i he ngaue he taimini.
 PRS in ART:INSPEC work ART:INSPEC time
 ‘She is at work right now.’
 Tongan (Abner & Burnett 2010: 5)
- (538) a. Na‘e fakamo‘ui [ha kilia] ‘a Sīsū.
 PST healthy.CAUS ART:NSPEC leper ERG Jesus
 ‘Jesus healed a leper.’
 b. Na‘a ne fakamālō‘i‘ai‘i ‘a Sīsū.
 PST 3SG CAUS.thankful ABS Jesus
 ‘He was thankful towards Jesus.’
 Tongan (Abner & Burnett 2010: 15)
- (539) Na‘e ma‘u ‘e Manu [ha pousikaati] mei hono kaume‘a nofo ‘i
 PST get ERG Manu ART:NSPEC postcard from 3SG.OBJ.GEN friend reside LOC
 Ha‘amoā mo [ha ngaahi pousikaati] mei hono kau kaainga nofo ‘i
 Samoa with ART:NSPEC several postcard from 3SG.OBJ.GEN PL relative reside LOC
 Hawai‘i.
 Hawaii
 ‘Manu received a postcard from her friend living in Samoa and several postcards from her
 relatives living in Hawaii.’
 Tongan (Ahn 2016: 12)
- (540) Na‘e ‘aka ‘e [ha faifekau] ‘a e tēvolo.
 PST kick ERG ART:NSPEC priest ABS ART:INSPEC devil
 ‘A priest kicked a devil.’
 Tongan (Abner & Burnett 2010: 9)

I maintain that *ha* is a nonspecific article at its core, even though it clearly can occur with specific referents in certain cases. To what extent its use with specific referents is subject to inter-speaker variation will have to be examined in more detail in future work. The important point for the present analysis is that its overall distribution does correspond to the contexts in which a non-specific article is expected, and we saw for certain contexts that *ha* is indeed incompatible with a specific interpretation of the referent. The occurrence of *ha* with specific referents seen in the last examples appears to be an extension of its main function and may be a more recent development.

An explanation of this extension involves the inclusive-specific article (*h*)e, which is the semantically appropriate article for marking specific referents. Especially the specific referents of *fefine poto* ‘wise woman’ and *kilia* ‘leper’ in (537) and (538) are discourse-prominent in that they

are first established in the discourse and then become the new center of attention in the following utterance. It is plausible that using the inclusive-specific article *he* is too ambiguous in such cases, as it does not allow to disambiguate between a definite and a specific interpretation. Due to their discourse-prominence, these referents are likely to be interpreted as identifiable. To ensure that this does not happen and that the hearer understands that the referent is not (yet) identifiable, or to signal that exact identity of the referent is not relevant or unknown, the speaker may resort to using *ha*. Although the semantic content of *ha* as a nonspecific marker is in conflict with the specificity of the referent, its pragmatic effect of signaling that the exact identity is irrelevant or not known (similarly to free-choice contexts) may outweigh the semantic mismatch in contexts in which the speaker wants to signal the irrelevance of identifiability. If this happens repeatedly and if such uses are conventionalized over time across the entire speaker community, *ha* may dissociate more and more from its semantic properties, which could lead to its development into an indefinite article. This is a rather hypothetical scenario at this point, since we saw examples like (526), (527), or (531), where *ha* marked the referent as necessarily nonspecific. In addition, examples like (523) showed that *ha* is incompatible with specific referent, at least in certain contexts (cf. Section 7.1.3 for similar examples including definite contexts). Therefore, I analyse *ha* as a nonspecific article.

6.4.4 On the source of nonspecific articles

This Section proposes a diachronic development from (verbal) irrealis marker to (nominal) nonspecific marker. There is evidence that this development has taken place in a number of languages in Central and North America in a few Mayan, Wakashan, and Siouan languages.

For instance, Q'anjob'al (Mayan) has a marker that may be a nonspecific article. Descriptions of the language usually mention *jun* as an indefinite article (Mateo Toledo 2017, Raymundo González et al. 2000, Zavala 1992), which is a cognate of the indefinite article in Tz'utujil presented in Section 6.1.1. However, consider examples (541) and (542). These examples suggest that we find a complementary distribution between *jun* and *junoq*, the former being used with specific and the latter with nonspecific referents.

- (541) a. ay mi [junoq / *jun yatut thioxh] b'ay yich calle?
 EXIST Q ART:NSPEC / ART:EXSPEC house.3SG God at back street
 'Is there a church at the end of the road?'
- b. ay [jun / *junoq yatut thioxh] b'ay yich calle.
 EXIST ART:EXSPEC / ART:NSPEC house.3SG God at back street
 'There is a church at the end of the road.'
- Q'anjob'al (primary data)

- (542) a. tzeb'ach yul [jun / *junoq tuktuk] mayal wawrtej naq
 come.IMP in ART:EXSPEC / ART:NSPEC tuktuk already called CL
 tz'umon ch'en.
 driver
 'Let's take a (certain) tuktuk. I already called the driver.'
- b. asi' yul [junoq / *jun tuktuk].
 go.IMP in ART:NSPEC / ART:EXSPEC tuktuk
 'Let's take a (any) tuktuk.'
 Q'anjob'al (primary data)

Two additional examples showing *junoq* with nonspecific referents are given in (543) and (544).

- (543) ma ay [junoq ch'en ha-tz'ib'al]?
 NEG EXIST ART:NSPEC CL:stone 2SG-pen
 'Do you have a(ny) pen?'
 literally: 'Don't you have a (any) pen?'
 Q'anjob'al (primary data)
- (544) ta' watz' max hot mulnaj hab'a ay mi wal [junoq / *jun
 if good PST COND work.2SG 2SG EXIST for.2SG EMPH ART:NSPEC / ART:EXSPEC
 ha-mimam na]
 2SG-big house
 'If you had worked well, you would have had a big house.'
 Q'anjob'al (primary data)

Zavala (1992: 137) provides similar contexts and notes that the marker *junoq* (spelled *xunox*) can be segmented into *jun* and the irrealis marker *-oq*. Two more examples of *junoq* are shown in (545) and (546).¹⁷ In (545), *junoq* is used within the scope of an intensional predicate. Example (546) is an instruction (probably to prepare food) and is non-episodic, which accounts for the use of *junoq*.

- (545) tši-w-otše [xunox tšeeal xii=an] katu? [xunox k'alan k'isis].
 INCOMPL-S:1.O:3-want ART:NSPEC CL oak=POSS and ART:NSPEC third cypress
 'I want a load of oak and a third of cypress.'
 Q'anjob'al (Zavala 1992: 206)¹⁸

¹⁷Martin (1977) provides some examples of Q'anjob'al *junoq* marking nonspecific referents in question and negation contexts as well.

¹⁸The original Spanish translation in Zavala (1992: 206) is: "Quiero una carga de roble y un tercio de ciprés".

- (546) tši-mey-le [xunox wotsan ʔaxonxoli] [xunox ʔonsa klavo]
 INCOMPL-toast-PASS ART:NONSPEC handful sesame ART:NONSPEC ounce clove
 [xunox-ša b'osan pimienta].
 ART:NONSPEC-other pinch pepper
 'You toast a handful of sesame, an ounce of cloves and another pinch of pepper.'
 Q'anjob'al (Zavala 1992: 205)¹⁹

Although neither Raymundo González et al. (2000) nor Mateo Toledo (2017) mention *junoq*, Mateo Toledo (2017) discusses *-oq* as a verbal irrealis or infinitive marker *-oq* for intransitive contexts. As suggested by Zavala (1992), it is very likely that this irrealis marker *-oq* fused with *jun* and developed into a nonspecific marker or article. Examples of the verbal uses of *-oq* are given in (547) to (549). They show its three uses as irrealis, future, and infinitive marker.

- (547) q-q'anjab' ayach ta q-ach q'anjab'-oq.
 POT-talk to.2 COND POT-2PL talk-IRR
 'She/he will talk to you, if you talk.'
 Q'anjob'al (Mateo Toledo 2017: 537)

- (548) hoq-ach txaj-l-oq yekal.
 POT-2 prayer-INTR-IRR tomorrow
 'You will pray tomorrow.'
 Q'anjob'al (Mateo Toledo 2017: 537)

- (549) x-toj heb' aw-j-oq b'ay-tu.
 IPFV-3 3PL voice-INTR-INF PREP-DEM
 'They went to shout there.'
 Q'anjob'al (Mateo Toledo 2017: 537)

Interestingly, Mateo Toledo (2017) provides two examples of *-oq* on nouns, shown in (550) and (551) below. In example (550), the irrealis marker occurs with a noun that functions as the predicate of the sentence. In (551), on the other hand, we see the marker on a noun that is in an argument position. It may have a nonspecific referent, as it is in the scope of the question.

- (550) man anima-oq hach.
 NEG person-IRR 2SG
 'You are not a person.'
 Q'anjob'al (Mateo Toledo 2017: 551)
- (551) maktxel max h-aq'-kan ko-taynomal-oq?
 who PFV 2SG-give-DIR:stay 1PL-guard-IRR
 'Who did you leave as our guard?'
 Q'anjob'al (Mateo Toledo 2017: 553)

¹⁹The original Spanish translation in Zavala (1992: 205) is: "Se tuesta un puñito de ajonjolí, una onza de clavo y otra pizca de pimienta".

These examples suggest that the irrealis marker is flexible enough to occur on nouns as well, even though it is not clear why it combines with the article *jun* in some cases with the noun in others. Determining whether this involves dialectal variation, and whether the use of *junoq* with nonspecific referents is systematic as the examples shown here suggest will require more research. The important insight from the Q'anjob'al patterns is that a verbal irrealis marker can be used to mark nonspecific referents.

In Ch'ol and Mocho, two other closely related Mayan languages spoken in Mexico, the use of the irrealis marker is attested on the indefinite article as well. Example (552) shows this for the irrealis marker *-ik* in Ch'ol. Note that in Ch'ol, we cannot speak of a nonspecific article, since the irrealis marker *-ik* is not used systematically in nominal nonspecific contexts.

- (552) ma'an [jum-p'ej=ik vivienda].
 EXIST.NEG ART:INDEF-CL=IRR residence
 'There isn't any residence.'
 Ch'ol (Vázquez Alvarez 2011: 239)

Martin (1998: 199) shows for Mocho that the irrealis marker *-oq* can be used with predicative nouns, e.g. in a negated context as in (553). However, she also notes that *-oq* can be used with the "indefinite determiner *huun-* 'one', where *-oq* replaces the definite numeral formative *-e'*". This is shown in (554). Although the translation of (554) does not suggest that *-oq* would mark nonspecificity, Martin (1998: 199) further remarks that "[s]uch constructions are generally glossed by *unos/unas* 'some' in local Spanish." Thus, the function of *-oq* in such cases may be to indicate the irrelevance of the exact identity of the referent, which was shown to be a common use of nonspecific articles in other languages as well.

- (553) ya w-eet-he muu aaw-ichmaal-oq-in.
 because 1SG-RELAT:SELF-PRT NEG 2SG-husband-IRR-1SG
 'Because, as for me, I'm not your husband.'
 Mocho (Martin 1998: 198)
- (554) heel-bi' ya cho x-k'uul-a' kene' [huun-oq x-hu'm-aal].
 if-PRT because for A:3-do-TR DIR.2:remain one-NONSPEC 3-paper-ABS
 '(It's good) if he leaves the papers complete.'
 Mocho (Martin 1998: 199)

Another language showing a similar pattern is Hidatsa, a Siouan language spoken in the USA. In contrast to the Q'anjob'al examples, the nonspecific interpretation of a nominal expression with the irrealis marker only appears to be additional to the irrealis predicate interpretation. The so-called conditional marker *-rug* in Hidatsa usually marks irrealis, conditional, and future on both verbs and nouns. Examples (555) and (556) illustrate its use with a verb and a noun, respectively.

(555) ééhgee-wa-**rúg** aru-nii-ma-hgi.wé'-he
 know-1.ACT-**COND** IRR-2.STAT-1.ACT-tell-EMPH
 'If I knew it I would tell you.'
 Hidatsa (Park 2012: 228)

(556) áàda-**rúg** magi-maa-íhgohbi-wihi-'o-'?
 daylight-**COND** RECIP-1.ACT-meet-FUT.Q.1-PL-Q
 'Shall we meet tomorrow?'
 Hidatsa (Park 2012: 181)

The Hidatsa article system consists of the definite article *-s* and the indefinite article *-wa* (Park 2012). Also for Hidatsa, it was noted that the conditional marker *-rúg* can attach to nouns and trigger a nonspecific interpretation of the referent (Park 2012: 367). An example of this is shown in (557), contrasted with the expression of a specific referent in (558):

(557) hiraacá-mià-**rúg** aru-m-úá-waa-c.
 Hidatsa-woman-**COND** IRR-1-marry-CAUS.1-DECL
 'I am going to marry a Hidatsa woman.'
 literally: 'If she is a Hidatsa woman I will marry her.'
 Hidatsa (Park 2012: 368)

(558) hiraacá-mià-**wa** m-úá-waa-c.
 Hidatsa-woman-**ART:INDEF** 1-marry-CAUS.1-DECL
 'I married a Hidatsa woman.'
 Hidatsa (Park 2012: 368)

Also in Crow, yet another Siouan language, the conditional marker *-dak/-lak* may be related to an indefinite marker used with nouns. Crow, like Lakota, has an article system with a definite, specific, and a nonspecific article.²⁰ However, as was mentioned in Section section 2.2.4, in a certain narrative genre, the exclusive-specific and the nonspecific articles are replaced by the indefinite marker *-dak/-lak* (Graczyk 2007: 233). The use of *-dak* is illustrated in examples (559) and (560) below; it no longer distinguishes between specific and nonspecific referents.

(559) bachee-**lák** baa-aash-dée-k.
 man-**ART:INDEF** INDEF-hunt-go-DECL
 'A man went hunting.'
 Crow (Graczyk 2007: 230)

²⁰The form *-dak* follows consonants while *-lak* is used after vowels.

6.5 Summary

This chapter discussed the article types found in the indefinite domain: indefinite, presentational, exclusive-specific, and nonspecific articles. After showing examples of indefinite articles in Section 6.1, I argued in Section 6.2 that we need to distinguish presentational articles as a distinct type of indefinite articles. On the one hand, these articles are restricted to the coding of discourse-prominent referents; on the other hand, they can also be used with nonspecific referents. This shows that the traditionally assumed semantic (from specific to nonspecific) and discourse-pragmatic (from prominent to less prominent) developments of indefinite articles may be correlated but do not necessarily depend on each other.

Section 6.3 presented exclusive-specific articles. They are similar to and yet different from presentational articles, since they only mark specific referents without being restricted to discourse-prominent referents. Nonspecific articles, the last type of articles in the indefinite domain, were presented in Section 6.4. Nonspecific articles can also be understood as the counterpart of exclusive-specific articles within the indefinite domain, being restricted to the occurrence with nonspecific referents. I further showed that verbal irrealis markers are a possible diachronic source for nonspecific articles and nonspecific markers in general. The first step in their development is their extension to nominal predicates in irrealis contexts, then to nouns with nonspecific referents in irrealis contexts, and eventually, they may become systematic nonspecific markers, i.e. nonspecific articles.

Chapter 7

Domain-crossing articles

This chapter presents examples of two article types that are semantically vague between referents from both the definite and indefinite domains. The first type is the inclusive specific article that is used to mark both definite and specific referents. Sections 7.1.1 to 7.1.3 present the inclusive-specific articles in Bemba, Ayoreo, and Tongan. As will be shown, the article in Tongan should better be analysed as a weak inclusive-specific article.

The other article type described in this chapter is the referential article (Section 7.2), marking definite, specific, and nonspecific referents. Referential articles thus indicate referentiality as such rather than a specific referential function. The three examples presented in more detail are the articles in Rapa Nui (Section 7.2.1), Baure (Section 7.2.2), and Halkomelem (Section 7.2.3).

Similarly to the articles presented in Chapter 6, inclusive-specific and referential articles interact in a different way with contexts of negation as in *I do not see a(ny)/the fish in the tank*. If the referent is marked by an indefinite article in English, the referent is interpreted as nonspecific. With an inclusive-specific article, such referents will be interpreted as definite by default, given that the article is incompatible with a nonspecific interpretation and given that a specific interpretation of the referent would be marked. Referential articles, on the other hand, are compatible with both referent types and are thus ambiguous between a definite or a nonspecific interpretation of the referent within the scope of negation, at least in the absence of further disambiguation from the context.

7.1 Inclusive-specific articles

In this section, I present three examples of inclusive-specific articles from Bemba, Ayoreo, and Tongan. Inclusive-specific articles co-express definite and specific referents and are thus semantically vague between these two referent types. Bemba only has an inclusive-specific article, whereas Ayoreo and Tongan have a nonspecific article as well. The Tongan system is even more complex in

that it also has an anaphoric article, making the inclusive-specific article a weak inclusive-specific article. Inclusive-specific articles are not frequently attested across the world's languages; my sample only contains 4 instances (including the weak article in Tongan).

7.1.1 Bemba

Bemba is a Narrow Bantu language spoken mainly in the north-east of Zambia, as well as in parts of the Democratic Republic of Congo, Malawi, and Botswana. Bemba has over 4 million speakers.¹

Nouns in Bemba, like nouns in various other Bantu languages (e.g. Zulu, Kirundi, Nguni), do not only have prefixes that indicate number and gender (traditionally referred to as noun classes). They also feature an additional prefix, often called “pre-prefix” or “augment”.² An example is given in (563):

- (563) *u-mu-ntu* *aaliishile*.
ART:INSPEC-CL1-man come.PST.3SG
'The/a man came.'
Bemba (Givón 1969: 47)

The noun *umuntu* ‘man’ in (563) consists of the lexical root *ntu*, a prefix *mu-* indicating that it belongs to class 1, and the additional prefix *u-*. I treat this prefix as the inclusive-specific article in Bemba. Table 7.1 gives an overview of the noun class prefixes and the corresponding article exponents based on Givón (1969: 28-31). The rightmost column indicates the class prefix as it occurs on nouns in the absence of the article; the column in the middle presents the combined forms (i.e. the complex noun prefix) consisting of the article and the noun class marker.

Classes 1 to 14 are singular and plural pairs of what could be treated as 7 different gender classes. Class 15 contains abstract or derived nouns and also infinitives. Classes 16 to 18 are somewhat different in that they are locative markers (*pa* ‘at’, *ku* ‘at/in’, and *mu* ‘inside’) which can be added to nouns of other classes (the original class marker is retained) to express spatial but also other abstract semantic values. As can be seen in Table 7.1, the augment or what I treat as an inclusive-specific article has different exponents according to the class of the noun.

While Hoch (1964) only lists morphosyntactic factors to account for the absence of the augment, I follow Givón (1969) in that it is conditioned by the referent type of the noun. To be precise, Givón (1969) uses a different terminology, distinguishing between specific (token) vs. generic (type) expressions. An example to illustrate the difference, motivating the labels used by Givón, is given in (564).

¹Cf. <https://www.ethnologue.com/language/bem>.

²The form but also the factors that condition the distribution of the augment differ significantly across Bantu languages. In some languages the use of the augment is tied to definiteness or topicality, in others it only occurs in certain syntactic contexts (e.g. is absent with negation), and in some languages nouns can no longer occur without the augment. For an overview of the the augment in Bantu languages, see de Blois (1970).

Table 7.1: Bemba noun class markers

noun class	ARTICLE-CLASS	CLASS
1	u-mu-	mu(u)-
2	a-ba-	ba(a)-
3	u-mu-	mu(u)-
4	i-mi-	mi(i)-
5	(il)i-	li(i)-
6	a-ma-	ma(a)-
7	i-ci-	ci(i)-
8	i-fi-	fi(i)-
9	i-N-	ni(N)-
10	i-N-	ni(N)-
11	u-lu-	lu(u)-
12	a-ka-	ka(a)-
13	u-tu-	tu(u)-
14	u-bu-	bu(u)-
15	u-ku-	ku(u)-
16	/	pa-
17	/	ku-
18	/	mu-

- (564) a. *imfumu ili u-mu-puupu.*
 chief.ART:INSPEC be.3SG ART:INSPEC-CL1-thief
 ‘The chief is a proven, known, thief.’
- b. *imfumu muu-puupu.*
 chief.ART:INSPEC CL1-thief
 The chief is by nature, inherently, a thief.
 Bemba (Givón 1969: 37)

In the construction in (564a), where *umupuupu* ‘thief’ includes the pre-prefix, the nominal expression is interpreted as a referential expression which is ambiguous between a definite and a specific reading. The construction thus is more like an identificational construction, and because the referent is at least semantically specific, it is called specific by Givón. In (564b), on the other hand, the noun *muupuupu* ‘thief’ occurs without the pre-prefix. Here, it can only be interpreted as a nominal predicate, i.e. it is not referential in this case. Givón (1969) labels this use generic. In the examples themselves, Givón (1969) marks the distinction of specific vs. generic as “token” vs. “type”, respectively. While “specific” is used in the same way here as in Givón (1969), I call contexts in which a nominal expression is not referential “non-referential”. This is also the terminology used in his later work (e.g. Givón 1984); where he distinguishes between referential and non-referential expressions in predicative vs. identificational constructions.

Since the forms with the pre-prefix are used to mark that the noun has a definite or specific referent, and since they cannot be used with nonspecific referents, I analyse the pre-prefix as an inclusive-specific article. If the ambiguity between interpreting the referent as definite or specific is not resolved by the context, the inclusive-specific article is always translated as *the/a* into English. This shows that the article does not disambiguate between a definite and an indefinite interpretation and thus co-expresses those referential functions.

Two examples to show this are given in (565) and (566). The inclusive-specific article *u-* is obligatory in specific or definite contexts (566a), the lack thereof being ungrammatical as in (566b):

(565) *naalimweene *(a)-ba-ana.*
 see.PST.1SG ART:INSPEC-CL2-child
 ‘I saw (some/the) children.’
 Bemba (Givón 1969: 42)

(566) a. *u-mu-ntu aaliishile.*
 ART:INSPEC-CL1-man come.PST.3SG
 ‘The/a man came.’
 b. **muu-ntu aaliishile.*
 CL1-man come.PST.3SG
 intended: ‘The/a man came.’
 Bemba (Givón 1969: 47)

As was mentioned in the beginning of this chapter, an argument within the scope of a negated single past event like *I didn’t see a/the child* can be interpreted as having either a definite or a nonspecific referent. Importantly, the referent cannot be specific. Having an inclusive-specific article but no article to encode nonspecific referents leads to the following effect in Bemba. If the inclusive-specific article is used, the referent is necessarily interpreted as definite, no longer being ambiguous between a definite and a specific reading. The absence of the article in such contexts marks that the referent is nonspecific. This contrast is shown in (567). In (567a), the referent of *muana* ‘child’ is necessarily definite and necessarily nonspecific in (567b). Example (568) shows the same contrast for the referents of *muntu* ‘man’ and *muana* ‘child’. Note that the inclusive-specific article is absent on the first of the two coordinated nouns; this will be discussed below.

(567) a. *nshiamwene u-mu-ana.*
 see.PST.NEG.1SG ART:INSPEC-CL1-child
 ‘I didn’t see the child.’
 b. *nshiamwene mu-ana.*
 see.PST.NEG.1SG CL1-child
 ‘I didn’t see any child.’
 Bemba (Givón 1969: 42)

- (568) a. nishiatumine fi-tabo ku-mu-ntu na-**u**-mu-ana.
 send.PST.NEG.1SG CL8-book CL17-CL1-man and-**ART:INSPEC**-CL1-child
 ‘I didn’t send any books to the man and the child.’
- b. nishiatumine fi-tabo ku-mu-ntu na-**mu**-ana.
 send.PST.NEG.1SG CL8-book CL17-CL1-man and-**CL1-child**
 ‘I didn’t send any books to any man and any child.’
- Bemba (Givón 1969: 56)

Thus, in general, the inclusive-specific article is required in definite as well as specific indefinite contexts and excluded in nonspecific contexts. Nevertheless, we find contexts like (568a) where the inclusive-specific article does not occur on all nouns with a definite referent (Givón 1969). The inclusive-specific article is blocked morphosyntactically in such contexts. Another similar example can be seen in example (569). In (569a), the use of the inclusive-specific article *-u* is again blocked on a nouns with the additional class 17 prefix, which marks it as the indirect object. We see that it is the presence of the additional class 17 prefix *ku-* that blocks the use of the inclusive-specific article when comparing (569a) to (569b). In (569b), the class 17 marker *ku* is only present on the first of two coordinated nouns, namely on *muntu* ‘man’. Being absent on the second noun *muana* ‘child’, the latter occurs with the inclusive-specific article *u-* as expected.

- (569) a. aamoneshya i-ci-tabo ku-(***u**)-mu-ana.
 show.PST.3SG ART:INSPEC-CL7-book CL17-**ART:INSPEC**-CL1-child
 ‘He showed the book to the child.’
- b. naalimoneshya i-ci-tabo ku-mu-ntu na-**u**-mu-ana.
 show.PST.1SG ART:INSPEC-CL7-book CL17-CL1-man and-**ART:INSPEC**-CL1-child
 ‘I showed the book to the man and the child.’
- Bemba (Givón 1969: 55f)

Other prefixal markers that block the use of the inclusive-specific article are a prenominal demonstrative as in (570), an associative marker shown in (571), or a locative marker (class 18 marker) as in (572).

- (570) uyu-(***u**)-mu-ntu
 DEM-**ART:INSPEC**-CL1-man
 ‘this man’
 Bemba (Givón 1969: 53)

- (571) u-mu-ana u-a (***u**)-mu-luungi
 ART:INSPEC-CL1-child CL1-AM **ART:INSPEC**-CL1-hunter
 ‘the child of the hunter’
 Bemba (Givón 1969: 54)

- (572) ali mu-(*u)-mushi
 be.3SG CL18-ART:INSPEC-village
 ‘He’s in the village.’
 Bemba (Givón 1969: 55)

7.1.2 Ayoreo

Ayoreo, whose nonspecific article was presented in Section 6.4.1, also has an inclusive-specific article that encodes definite and specific referents. The present section will briefly present the Ayoreo inclusive-specific article. Like the nonspecific article, the inclusive-specific article is realized as an affix on the noun with different exponents according to the number and gender value of the noun. The exponents are repeated in Table 7.2 for convenience.

Table 7.2: Ayoreo article exponents

		inclusive-specific	nonspecific
FEM	SG	-a/-ia/-∅	-raque/-taque/-naque
	PL	-die/-i	-rigi/-tigi/-niŋi
MASC	SG	-i	-tique/-rique/-nique
	PL	-ode	-tigo/-rigo/-niŋo

The following examples show the inclusive-specific article in various definite contexts. In (573b), we see the anaphoric referent of *ibocadie* ‘their rifles’, mentioned for the first time in (573a).³ The anaphoric referent is marked by the inclusive-specific article. Note that when the referent is mentioned for the first time in (573a), it is equally marked by *-die*.

- (573) a. Mu ujeta ore ch-amurase **da-boca-die** nga que
 but COMP 3PL 3-put.down 3.REFL-fire.weapon-ART:INSPEC.F.PL COORD NEG
 pit-ode ore.
 fear-ART:INSPEC.M.PL 3PL
 ‘But in order for them to put down their own rifles, they should not have fear.’
- b. Yoqui-todo-die u nanique ore i-boca-**die**.
 1PL-fears-ART:INSPEC.F.PL COP time.back 3PL 3-fire.weapon-ART:INSPEC.F.PL
 ‘In those days, **their rifles** were our (cause for) fear.’
 Ayoreo (Bertinetto 2009: 54)

³The examples taken from Bertinetto (2009) do not segment the nominal stem and the article. I added this segmentation to the examples based on the extensive morphological descriptions and paradigms given in Ciucci (2016).

In (578) and (579), the referents marked by the inclusive-specific article occur in the position of the direct object.

(578) Ore ch-udute [dara quedéjn-ane].
 3PL 3-hear voice.M different-ART:INSPEC.M.SG
 ‘They heard strange voices.’
 Ayoreo (Ciucci 2016: 486)

(579) Ch-imo [carataque gare queújn-ane] iji ta.
 3-see jaguar.M two very.big-ART:INSPEC.M.PL LOC there
 ‘He saw two big jaguars right there.’
 Ayoreo (Ciucci 2016: 486)

Example (580) shows that a noun marked by the inclusive-specific article can also be interpreted as specific in the subject position.

(580) Chuningame mu eti uyujna-i chubuchu aja quedejn-ane.
 gets.suprised but COMP storm-ART:INSPEC.SG blows LOC different-ART:INSPEC.M.PL
 ‘All of a sudden, a storm blew in every direction.’
 Ayoreo (Ciucci 2016: 46)

That the inclusive-specific article cannot be used to mark nonspecific referents was already shown in Section 6.4.1. To show that the inclusive-specific article is in fact a referential marker and not the default form of the noun occurring in all types of contexts but nonspecific ones, the following examples show that the article is not used in predicative contexts. Example (581) shows a minimal pair for the nominal expression *arócojna-quedejna* ‘alligator’. It is first used predicatively, and we see that it occurs without the article. The second use is referential, and indeed the noun is marked by the inclusive-specific article *-i*. Another example of bare nouns used as predicates is shown in (582).

(581) ¡Cajire to! ¡Arócojna-quedejna! ¡Arócojna-quedejna-i deji ne!
 look too alligator-different.M.SG alligator-different-ART:INSPEC.M.SG EXIST.3 there
 ‘Look there! It is an alligator! There is an alligator right there!’
 Ayoreo (Ciucci 2016: 485)

(582) Que yame go. Que cuchiso go. Que uñec u=po go.
 NEG monkey.F POL NEG animal.M POL NEG other COP=also POL
 ‘It was no monkey, no animal, nothing of the sort.’
 Ayoreo (Ciucci 2016: 485)

To conclude, this Section showed that what Bertinetto (2009) and Ciucci (2016) call the base form of the Ayoreo noun is in fact an inclusive-specific article. It systematically marks that the noun expresses a definite or a specific referent. It does not mark nonspecific referents and cannot be used in predicative, i.e. non-referential contexts either.

7.1.3 Tongan: Evidence for a weak inclusive-specific article?

As was mentioned in Section 6.4.3, Tongan (Oceanic, Tonga) has a system of three articles: an anaphoric, a (weak) inclusive-specific, and a nonspecific article.⁵ This section focuses on the former two articles, showing that *(h)e* could be analysed as a weak inclusive-specific, and its combination with the so-called “definite accent” (DA) as an anaphoric article. The DA goes back to Churchward (1953), who introduces it as a stress shift towards the right edge of the nominal complex.⁶ According to MacDonald (2014: 16-18), this distinct stress pattern signals anaphoric referents, and also Hendrick (2005: 918) notes that the DA is required for a referent to be marked as anaphoric. If this is indeed the case, then the combination of *(h)e* and the DA should be analysed as an anaphoric article, whereas *(h)e* alone without the DA is a weak inclusive-specific article by analogy to weak definites articles.

The inclusive-specific article is realized as *he* when it immediately follows the prepositions *e*, *i*, *ki*, and *mei*; it is realized as *e* in all other contexts (Völkel 2010: 159). For reasons of simplicity, I will refer to the inclusive-specific article as *he* in the text. That it occurs with definite referents is shown in (583). Here, the referent of *liliú* ‘change’ is anaphoric and refers back to a previously mentioned referent. As expected, the second mention of *liliú* ‘change’ in (583) occurs with the *he* and the DA. Note that the first mention of *liliu* ‘change’ is marked by the nonspecific article *ha*.

- (583) Ko e tokotaha kotoa pē ‘oku pole ke ne fakahoko [ha
 PRT ART:INSPEC individual while whatever PRES challenge to 3SG implement ART:NSPEC
liliu] ‘o ha fa‘ahinga Tui, Tokateline pē Fa‘unga Faka-Politikale kuopau
 change of ART:NSPEC religion doctrine or structure political PERF.certain
 ke ne ma‘u ‘a e ngaahi nāunau ‘o [e feinga *liliú*].
 PRT 3SG get ABS ART:INSPEC PL resource of ART:INSPEC try change.ART:ANA
 ‘Anyone who dares to change a religion, doctrine, or political structure must have the
 means to make that change’
 Tongan (Hendrick 2005: 915)

Another example of an anaphoric referent marked by *he* is given in (584). Hendrick (2005: 915) notes that *he* together with the anaphoric article can be used with intensional predicates, where the referent, *fefine potó* ‘clever woman’ needs to be interpreted as identifiable due to previous mention.

⁵As in Section 6.4.3, I will only focus on the neutral forms here and ignore the so-called “emotional” or diminutive article forms.

⁶The phonetic and prosodic details of the definite accent go beyond the scope of this study. It is important to note, though, that the stress shift of the definite accent has been analysed as the result of vowel lengthening (cf. R. Clark 1974, Schütz 2001, Anderson & Otsuka 2006, Garellek & White 2015, and references therein). I thank Svenja Völkel for pointing this out to me.

(584) 'Oku kumi 'a Siaosi ki [he fefine potó].
 PRES look ABS Siaosi for ART:INSPEC woman clever.ART:ANA
 'Siaosi is looking for the clever woman (previously talked about).'
 Tongan (Hendrick 2005: 915)

In (585), on the other hand, the inclusive-specific article *he* is used without the anaphoric article. In this case, the referent of *fefine potó* 'clever woman' is interpreted as specific indefinite.

(585) 'Oku kumi 'a Siaosi ki [he fefine potó].
 PRES look ABS Siaosi for ART:INSPEC woman clever
 'Siaosi is looking for a [certain] clever woman.'
 Tongan (Hendrick 2005: 914)

Also in establishing contexts, which correspond to another type of pragmatic definites, the inclusive-specific article *he* is used together with the anaphoric article. This is shown in (586).

(586) Ko [e me'alele kulokula na'e fakatau 'e Sioné].
 PRED ART:INSPEC car red PST buy ERG Sione.ART:ANA
 'It's the red car that Sione bought last year.'
 Tongan (Anderson & Otsuka 2006: 24)

Interestingly, Ahn (2016) notes that *he* alone can also be used with restrictive relative clauses. This is shown in (587b); here, the article *he* appears to have an establishing function. The use of *he* without the DA in this context suggests that it may be able to encode pragmatic definite functions at least in certain types of contexts.

- (587) a. Na'e ma'u 'e Manu ha pousikaati mei hono kaume'a nofo 'i
 PST get ERG Manu ART:NSPEC postcard from O:3SG.GEN friend reside LOC
 Ha'amoā mo ha ngaahi pousikaati mei hono kau kaainga nofo 'i
 Samoa with ART:NSPEC several postcard from O:3SG.GEN PL relative reside LOC
 Hawai'i.
 Hawaii
 'Manu received a postcard from her friend living in Samoa and several postcards from their relatives living in Hawaii.'
- b. Na'a ne fakapuliki 'a [e pousikaati na'e 'omai 'e hono kaume'a
 PST 3SG lose ABS ART:INSPEC postcard PST give ERG O:3SG.GEN friend
 nofo 'i Ha'amoā].
 reside LOC Samoa
 'She lost the postcard that her friend living in Samoa sent.'
 Tongan (Ahn 2016: 12)

Example (588) illustrates the use of the inclusive-specific article *he* with situationally unique referents. It marks the referent of *fonua* 'land' as unambiguously identifiable because of its salient

status in the immediate discourse situation. Interestingly, also a situationally unique referent is marked by both *he* and the anaphoric article. The referent of *vakapuna* ‘airplane’, on the other hand, is only marked by the inclusive-specific article *he*. In this context, it may in fact not be a situationally unique referent but be part of the lexical expression for ‘flying in an airplane’.

- (588) na‘a ku sio hifo ki [he fonuá] lolotonga ‘eku puna [he
 PST 1SG see down all ART:INSPEC land.ART:ANA while POSS:1SG fly ART:INSPEC
 vakapuna].
 airplane
 ‘I looked down to the land while I was flying.’
 Tongan (Völkel 2010: 117)

Another example with either an anaphoric or situationally unique (or potentially recognitional) referent is shown in (589). Again, the unambiguously identifiable referent of *fale mahakí* ‘hospital’ occurs together with the inclusive-specific and the anaphoric article. If the anaphoric article is not used, the referent is preferably interpreted as a specific referent, as can be seen in (590).

- (589) Ko Mele ‘oku ne ngaane ‘i [he fale mahakí].
 PRED Mele PRES 3SG work DAT ART:INSPEC house sick.ART:ANA
 ‘Mele works in the hospital.’
 (the only one possible, based on shared knowledge, or the one we have been speaking
 about)
 Tongan (MacDonald 2014: 53)

- (590) Ko Mele ‘oku ne ngaane ‘i [he fale mahaki].
 PRED Mele PRES 3SG work DAT ART:INSPEC house sick
 ‘Mele works in a hospital.’
 Tongan (MacDonald 2014: 53)

MacDonald (2014: 52) explains the circumstances of (589) and (590) as follows:

“Similarly, asked how to say, ‘Mele works in a/the hospital,’ my consultant indicates that *he fale mahaki* (without DA) (47a) [(589), LB] is generally preferred. *He fale mahakí* (with DA) (47b) [(590), LB] is only preferred when the hospital has been mentioned in the current discourse, although she considers it acceptable (but dispreferred) in contexts where it has not been mentioned if the speaker and hearer both know that there is only one hospital.”
 (MacDonald 2014: 52)

It is unclear to what extent this means that the inclusive-specific article alone can(not) mark the referent of *fale mahaki* as definite in (590). It does seem to be the case that the interpretation of the referent as specific is strongly preferred in that context.

Abner & Burnett (2010) provide an interesting minimal pair context for the inclusive-specific article and its combination with the anaphoric article, which probably also relates to situational uniqueness. Referring to the example shown in (591), they note that the anaphoric article (*e ... tu‘á*) can only be used if there is a devil in the context of the discourse or “if there is an actual devil that the speaker is pointing to in the context” (Abner & Burnett 2010: 21). Again, if the referent is interpreted as situationally unique, the anaphoric article must be used with the inclusive-specific article. Interestingly, Abner & Burnett (2010: 20) note that if there is no actual devil in the discourse situation, then the use of the anaphoric article is infelicitous and only the inclusive-specific article (*e ... tu‘á*) can be used.

- (591) Ko Piula, ‘oku tuli ‘a [e tēvolo ‘oku ne tui ‘oku ‘i tu‘a /
 PRED Pilua PRES chase ABS ART:INSPEC devil PRES 3SG believe PRES in outside /
 tu‘á].
 outside.ART:ANA
 ‘Pilua, she’s chasing this devil that she believes is outside.’
 Tongan (Abner & Burnett 2010: 21)

More evidence for *he* being a weak inclusive-specific article comes from example (592), taken from Anderson & Otsuka (2006: 23). Here, the inclusive-specific article *he* is not sufficient to mark the referent as unique in the discourse situation; instead, it can only be interpreted as marking it as non-identifiable to the speaker, i.e. as indefinite specific.

- (592) a. Ko e hā ē?
 PRED ART:INSPEC what that
 ‘What is that?’
 b. Ko [e me‘alele].
 PRED ART:INSPEC car
 ‘It’s a / *the car.’
 Tongan (Anderson & Otsuka 2006: 23)

So far, we only saw examples of *he* alone in which it was translated by an indefinite article in English and in which it was analysed as marking the referent as specific. Example (593) shows that *he* is also used with contextually unique referents. Here, it indicates that the referent of *taimi si‘i taha* ‘smallest time’ is unique within a certain context due to the presence of the superlative.

- (593) Ko [e taimi si‘i taha] kete ngaue totonu‘aki ‘a e NRT
 PRT ART:INSPEC time small most to work correctly ABS ART:INSPEC NRT
 koe uike 7.
 PRT.ART:INSPEC week 7
 ‘The smallest time for the NRT [Nicotine Replacement Therapy] to work correctly is seven weeks.’
 Tongan (Hendrick 2005: 911)

Also Otsuka (2000: 50) mentions that the use of the article *he* can mark the referent as definite or specific without disambiguating between these two referent types. This is shown in (594). Without further contextual information, the referent of *ta'ahine* 'girl', marked by the article *he* can either be identifiable by both the speaker and the hearer and thus definite, or, it can be a single particular but not identifiable referent, i.e. a specific referent.

- (594) Na'e sio 'a Sione ki [he ta'ahine].
 PST see ABS Sione to ART:INSPEC girl
 'Sione saw a/the girl.'
 Tongan (Otsuka 2000: 50)

Having seen that *he* is able to mark a referent as definite, the following two examples show its use with specific referents. Example (595) is a typical context where a new discourse referent is introduced; we see that *he* is used in such cases. In (596), the referent of *'ofa* 'present' is not relevant or not identifiable by the hearer and is therefore specific, and it is marked by *he* as well. As expected, *he* does not occur with the DA in such contexts.

- (595) Na'e 'i ai [e ongo mātu'a ko Tafi mo Ongo'alupe].
 PST in there ART:INSPEC DU parents PRT Tafi and Ongo'alupe
 'Once there was a couple named Tafi and Ongo'alupe.'
 Tongan (Hendrick 2005: 913)

- (596) ko 'eku lele mai pe 'o fakaa'u 'a [e 'ofa] mei he
 PRES POSS:1SG POLIT:run to just and POLIT:present ABS ART:INSPEC present ABL ART
 kāinga mei he Lolo 'a Halaevalu 'i teu fakamanatu 'a e
 family ABL ART:INSPEC Lolo 'a Halaevalu prepare remind ABS ART:INSPEC
 valu-ngofulu-ma-nima ta'u 'a e 'Afio na.
 85th year POSS ART:INSPEC POLIT:you
 'I have come with the relatives from Lolo a Halaevalu to present to your Majesty a gift for
 your eighty-fifth birthday.'
 Tongan (Völkel 2010: 208)

That *he* is used with specific referents while the DA, i.e. the anaphoric article, cannot occur in such contexts is explicitly shown in (597). For a specific interpretation of the referent of *puaka* 'pig' in an existential construction, only *he* can be used; the use of the anaphoric article as in (597b) being infelicitous.

- (597) a. 'Oku 'i ai [e puaka] 'i Māketi.
 PRES DAT there ART:INSPEC pig DAT Market
 'There is a pig at the Market.'

- b. 'Oku 'i ai [e *puaká] 'i Māketi.
 PRES DAT there ART:INSPEC pig.ART:ANA DAT Market
 intended: 'There is a pig at the Market.'

Tongan (MacDonald 2014)

In Section 6.4.3, we saw that the nonspecific article is used as a predicate marker in constructions with non-verbal predicates. Example (598a), the counterpart of (532), shows that the inclusive-specific article can occur in such contexts as well. If the anaphoric article is used as in (598b), we see that the construction is interpreted as an identificational construction. Here, *faiakó* 'teacher' is a referential expression and no longer a nominal predicate, which is reflected in the English translation by the use of the definite article.

- (598) a. Ko [e faiako] ia.
 PRT ART:INSPEC teacher he
 'He is a teacher.'
- b. Ko [e faiakó] ia.
 PRT ART:INSPEC teacher.ART:ANA he
 'He is the teacher.'

Tongan (Hendrick 2005: 911)

We also saw in Section 6.4.3 that the nonspecific article is used as a marker for secondary predicates. Example (599), the counterpart of (533), shows that the inclusive-specific article can occur in such contexts as well. In this case, the expression *taki* 'leader' seems to be referential, though; Hendrick (2005: 911) notes that if *he* is used, the expression is "uniquely instantiated". Contrary to simple predicate constructions, example (599b) shows that the anaphoric article cannot occur in this context.

- (599) a. 'Oku tau lau ia ko [he taki].
 PRES we regard him PRT ART:NSPEC leader
 'We regard him as a leader.'
- b. 'Oku tau lau ia ko [he *taki].
 PRES we regard him PRT ART:NSPEC leader.ART:ANA
 intended: 'We regard him as a leader.'

Tongan (Hendrick 2005: 911)

In order to be an inclusive-specific article, *he* must not occur in nonspecific contexts. As was already shown in section 6.4.3, Tongan uses a distinct nonspecific article in such contexts. Therefore, we can conclude that the article *he* is an inclusive-specific article: it systematically occurs with definite and specific referent types and it is absent in nonspecific contexts.

Its status as a weak inclusive-specific article, whose use is restricted to semantic definites in the definite domain, is somewhat less clear. The distribution of *he* and its co-occurrence pattern

with the definite accent, i.e. the anaphoric article, strongly suggests that *he* alone is preferred in uniqueness-based definite contexts, as opposed to its combination with the DA in pragmatic definite contexts, as was shown for anaphoric contexts in (583), (584), and for establishing contexts in (586). In addition, we saw that the combination of the inclusive-specific article and the DA be used to mark situationally unique referents in (588), (591), and (592). On the other hand, examples (591), (593), (523) showed that *he* alone is able to encode a referent as definite when contextually unique. If the distribution of *he* indeed is as the examples shown in this section suggests, *he* would be a weak inclusive-specific article, by analogy to weak definite articles discussed in Section 5.3. Weak definite articles being already extremely rare crosslinguistically, weak inclusive-specific articles are even rarer. To the best of my knowledge, Tongan *he* is the only example of this article type. Future research is required to uncover other examples of weak inclusive-specific articles and to shed more light on their crosslinguistic distribution.

7.2 Referential articles

This section discusses referential articles. Referential articles coexpress the three major referential functions: definite, specific, and nonspecific. Thus, they do not distinguish between referential functions but rather signal that a nominal expression is a referring expression. I present three examples of referential articles, namely the ones of Rapa Nui (Section 7.2.1), in Halkomelem (Section 7.2.3), and in Baure (Section 7.2.2). The referential articles in Rapa Nui and Baure are more evident examples, as they have a single exponent that is used with all three major referent types. The referential articles in Halkomelem, on the other hand, has different exponents that have been analysed as separate articles/determiners in the literature. This is due to a number of additional factors that operate along with referentiality and condition the use of the markers. However, in this section, I argue that they should be analysed as different exponents of a single referential article.

7.2.1 Rapa Nui

Rapa Nui is an Oceanic language spoken on Easter Island in Chile. The language has approximately 1000 speakers (Kieviet 2017a: 1). There are a number of markers that form the determiner system of Rapa Nui (cf. Kieviet 2017a: 232-246, Kieviet 2017b). As I will show in this section, the marker *te* is systematically used to mark definite, specific, and nonspecific referents and should be analysed as a referential article. Like many Oceanic languages, Rapa Nui also has a determiner, *a*, that is used to mark proper nouns. I will not discuss *a* in this section. Here, I focus on the referential article *te* and its relation to the predicate marker *he*.

Examples (600) to (603) show the article *te* in different types of definite contexts. In (600), *te* is shown with an anaphoric referent.

- (600) a. He ma'u he oho mai he tu'u, he pu'a he haka kōpiro mo te taura
 NTR carry NTR go hither NTR arrive NTR cover NTR CAUS ferment for ART:REF rope
 mo hiri.
 for braid
 'He carried them (mulberry and hauhau) away and covered them (with water) to ferment to braid a rope from them.'
- b. ...He totoi he oho mai i [te taura], he tu'u mai he here ki ruŋa
 ...NTR drag NTR go hither ACC ART:REF rope NTR arrive hither NTR tie to above
 ki te puku mā'ea ena, 'ā ka ŋōŋō rō ka ŋōŋō rō, 'ā ka
 to ART:REF boulder stone MED until CNTG firm EMPH CNTG firm EMPH until CNTG
 harara rō.
 stiff EMPH
 '...He dragged the rope and tied it to a stone boulder, so it was very firmly tied and the rope was taut.'
- Rapa Nui (Kieviet 2017a: 581-582)

Example (601) shows a contextually unique referent. The preceding discourse segment introduces a family. It sets the scene of the children going out and bringing food to the father who is not at home but at work. The children being on their way, the referent of *ara* 'road' is the only salient referent of its kind and is therefore unambiguously identifiable by all discourse participants. Again, the referent in question is marked by *te*.

- (601) He e'a ia tū ŋā poki era a te vāeŋa o [te ara] he ha'ere he
 NTR go.out then DEM PL child DIST by ART:REF middle of ART:REF road NTR walk NTR
 iri.
 ascend
 'The children went out by the middle of the road and walked up.'
- Rapa Nui (Kieviet 2017a: 576)

In (602), the referent of *nu'u* 'people' is deictic and situationally unique since the referent corresponds to the people that are present in the discourse situation.

- (602) ki a kōrua ki [te nu'u] hakaroŋo mai 'i a au he kī atu ...
 to PROP 2PL to ART:REF people listen hither IMM PROP 1SG NTR say away ...
 'To you, to the people listening, I tell you ...'
- Rapa Nui (Kieviet 2017a: 46)

Example (603) shows an establishing referent, whose identity is established in the relative clause following the noun *me'e* 'thing'. Again, we see that *me'e* 'thing' is marked by the article *te*.

- (603) *matu ki oho tatou ki Vērene ki u'i i [te me'e] haka 'ite mai*
 come.on HORT go 1PL.IN to Bethlehem to look ACC ART:REF thing CAUS know hither
ena.
 MED
 'Come, let's go to Bethlehem, to see the thing announced (to us).'
 Rapa Nui (Kieviet 2017a: 46)

Having shown that the article *te* systematically occurs in various definite contexts, we now turn to its occurrence in specific contexts. This is shown in (604) and (605). In both examples, the nouns are linked to particular referents that are not identifiable by the hearer. Example (604) shows a typical context in which a new discourse referent is introduced, while the exact identity of the specific referent of *henua e hitu* 'seven islands' does not appear to be relevant in (605).

- (604) *'i te noho iŋa tuai era 'ā [te tanjata e tahi] te 'iŋoa*
 at ART:REF stay NMLZ ancient DIST IDENT ART:REF man NUM one ART:REF name
ko Tu'uhakararo.
 PROM Tu'uhakararo
 'In the old times (there was) a man called Tu'uhakararo.'
 Rapa Nui (Kieviet 2017a: 238)

- (605) *ko tu'u 'ana a au ki ruŋa i [te henua e hitu].*
 PERF arrive CONT PROP 1SG to above at ART:REF land NUM seven
 '(In my dream) I arrived on seven islands.'
 Rapa Nui (Kieviet 2017a: 238)

In order to be analysed as a referential article, *te* is required to mark nonspecific referents as well. Examples (606) to (607) show that this is what we find in Rapa Nui. In (606) and (607), the nonspecific referent occurs within the context of a question scoping over the existence of the referents of *me'e* 'thing' and *ika* 'fish'. Example (608) features an intensional predicate which makes the referent of *pipi* 'shell' nonspecific by default. The same applies to the referent of *vai* 'water' in (609). Kieviet (2017a: 238-239) explicitly notes that *te* occurs with nonspecific referents in these examples. However, the referents are not marked as such by the use of the referential article. Rather, it is the context that forces a nonspecific interpretation, and the referential article *te* is rather compatible with the nonspecific interpretation in these cases.⁷

- (606) *e ai rō 'ana hō [te me'e] mo ta'e rova'a e te 'Atua mo aŋa.*
 IPFV exist EMPH CONT DUB ART:REF thing for NEG obtain AG ART:REF God for do
 'Would there be any thing that God is not able to do?'
 Rapa Nui (Kieviet 2017a: 192)

⁷Kieviet (2017a: 239) comes to a similar conclusion, namely that the article *te* does not mark definiteness or specificity as such.

(607) e ai rō ‘ā [te ika] o roto?
 IPFV exist EMPH CONT **ART:REF** fish of inside
 ‘Are there any fish inside?’
 Rapa Nui (Kieviet 2017a: 241)

(608) he kī ō’oku ki kō’ku ŋā poki taina era mo oho o mātou mo kimi i
 NTR say O.POSS:1SG to O.POSS:1SG PL child sibling DIST for go of 1PLEX for search ACC
 [te pipi].
ART:REF shell
 ‘I told my brothers and sisters that we would go to look for shells.’
 Rapa Nui (Kieviet 2017a: 239)

(609) ko mate atu ‘ana ki [te vai] mo unu.
 PERF die away CONT to **ART:REF** water for drink
 ‘I’m dying for water to drink.’
 Rapa Nui (Kieviet 2017a: 237)

Given that *te* occurs with definite, specific, and nonspecific referents, its status as a referential article is not uncontroversial. Based on the previous examples illustrating its use, one may be tempted to argue that it should rather be analysed as an argument marker or as a syntactic or nominal marker of some sort. The remainder of this section will address this question.

From a syntactic point of view, *te* occurs in complementary distribution with the marker *he*. The latter is primarily used to mark nouns used as predicates, which is why Weber (2003) and Kieviet (2017a: 242) analyse it as a predicate marker.⁸ The use of *he* with a nominal predicate is shown in (610). Moreover, we see in (611) that *he* cannot be used in argument positions.

(610) [he tajata] tau manu era.
PRED man DEM bird DIST
 ‘That bird was a human being.’
 Rapa Nui (Kieviet 2017a: 242)

(611) ko tike’a ’ā a au (i) [*he honu].
 PERF see CONT PROP 1SG ACC **PRED** turtle
 intended: ‘I have seen a turtle’
 Rapa Nui (Kieviet 2017a: 237)

This makes the analysis of *te* and *he* as syntactic markers even more appealing; one could argue that *he* is a predicate marker and *te* an argument marker. This is a plausible analysis especially since Rapa Nui has flexible word classes, meaning that many lexemes can be used as either nouns

⁸There is also a so-called ‘neutral aspect marker’ *he* in Rapa Nui, glossed as NTR. Kieviet (2017a: 316) notes: “*He* is the most common aspect marker. It probably developed from the nominal predicate marker *he*.”

or verbs. Example (612) illustrates this for the lexeme *poki* ‘(be) child’, which can be used as an argument as in (612a) or as a predicate as in (612b).

- (612) a. he pōrekoreko [te ŋā poki] ’i Tāhai.
 NTR born:red ART:REF PL child at Tahai
 ‘Children were born in Tahai.’
 Rapa Nui (Kieviet 2017a: 76)
- b. mai te hora era ō’oku e poki nō ’ana ...
 from ART:REF time DIST POSS.3SG.O IPFV child just CONT ...
 ‘From the time when I was a child ...’
 Rapa Nui (Kieviet 2017a: 76)

The article *te* certainly being an important indicator that *poki* ‘child’ in (612a) corresponds to a noun and not to a verb, also its position in the clause and e.g. number marking indicate the syntactic status of *poki*. I would argue that the analysis of *te* as an argument marker is not incompatible with its analysis as a referential marker or article. In fact, I regard the syntactic function of *te* as the result of its referential functions. Marking expressions as arguments or nouns in a syntactic sense is the consequence of marking lexemes as referential expressions as opposed to event-referring expressions, i.e. predicates.

The analysis of *te* as a referential article also accounts for other contexts in which *te* does or does not occur in a straight-forward way. Contexts where the article *te* cannot be used are:⁹ following the instrumental preposition *hai* ‘with’, following the comparative preposition *pē* ‘like’, and in appositions. These three types of contexts are shown in (613), (614), and (615), respectively. While no other pronominal marker occurs after the preposition *hai* in (613), examples (614) and (615) show that the use of the predicate marker *he* is required with *pē* and possible with nouns in appositions.

- (613) he tunu mā’ea vera haka hopu i te poki [hai vai vera].
 NTR cook stone hot CAUS bathe ACC ART:REF child INSTR water hot
 ‘He cooked (the water) with hot rocks, and bathed the child with hot water.’
 Rapa Nui (Kieviet 2017a: 264)
- (614) ’iti’iti [pē he kio’e] hāpa’o i te hare [pē he paiheŋa] haka ’āriŋa.
 small:red like PRED rat care.for ACC ART:REF house like PRED dog CAUS face
 ‘Small like a mouse, guarding the house like an insolent dog.’
 Rapa Nui (Kieviet 2017a: 244)

⁹For other minor constructions that condition the distribution of *te* and *he*, see Kieviet (2017a: 235-236).

- (615) he kī [e te matu'a tane era o Te Rau [(he) taŋata pū'oko o te nu'u
 NTR say AG ART:REF parent male DIST of Te Rau PRED man head of ART:REF people
 o Kapiti]] ...
 of Kapiti ...
 'The father of Te Rau, the leader of the people of Kapiti, said ...'
 Rapa Nui (Kieviet 2017a: 243)

The absence of *te* in these contexts can be accounted for by the conventionalization of *te* marking referential expressions. Especially for nouns that express instruments or that occur in similitive constructions, we can assume that they are typically not used to refer but rather to evoke a certain concept. In appositions, the referent is usually identical to the referent of its anchor (Heringa 2011: 5), which also motivates the absence of the article as a referential marker in such constructions. If such functionally motivated trends conventionalize, they can become part of the grammar, resulting in the incompatibility of the article with certain syntactic contexts. While *te* cannot occur with the prepositions shown in (613) and (614), the use of *te* is by no means generally blocked in the presence of prepositions. As we can see in (616), *te* occurs with the spatial preposition *i* 'at'. Such spatial expressions are typically referential, accounting for the use of the referential article in such cases.

- (616) he eke ki ruŋa [i te mā'ea e tahi].
 NTR go.up to above at ART:REF stone NUM one
 'He climbed on a stone.'
 Rapa Nui (Kieviet 2017a: 247)

A syntactic account of *te* generally captures its distribution as an argument marker as opposed to the predicate marker *he*. However, this would not account for the use of *te* with certain prepositions and its absence with others. Although I analyse *te* as a referential article for the purposes of the present study, I do not exclude its syntactic function as an argument marker, which should rather be viewed as a consequence of its referential functions and its interplay with the predicate marker *he*.

7.2.2 Baure

Baure is an Arawakan language spoken in northeast Bolivia. Danielsen (2007: 12) counts 55 speakers of Baure in 2007; the language is generally no longer spoken in every day conversations.

Baure has an article, *to/ti*, which I analyse as a referential article. The form *to* is used for nominal and plural nouns; feminine singular nouns can take either *to* or *ti*. For the sake of simplicity, I will refer to the article as *to* in the text. According to Danielsen (2007: 312), *to* has the following properties:

“The article does not mark definiteness. It is generally used with proper names and heavenly bodies, such as *to ses* ‘the sun’ or *to kihēr* ‘the moon’, and nouns that refer to weather phenomena, such as *to vir* ‘the wind’ or *to sowon* ‘the rain’. The article does not introduce characters in a narration. [...] The article can then be used when the NP is already topic and known to the hearer. The form *to* is the least marked or emphatic form of the determiners. It occurs more frequently in specific constructions, such as cleft or relative clauses and before nominalized verbs.”

Indeed, the examples provided in Danielsen (2007) suggest that other demonstratives are preferred over *to* to introduce new discourse-prominent referents.¹⁰ However, I will show in this section that *to* is systematically used with definite, specific, and nonspecific referents. Therefore, I argue that *to* is a referential article, marking that the nominal expression it occurs with is referential. Implicitly, Danielsen (2007: 310) comes to a similar conclusion; she notes that nominal expressions in Baure require the presence of a determiner (most often *to*) to be referential.

Examples of its use in definite contexts are given in (617), (618), and (619). In (617), *to* is shown with an anaphoric referent. The referent of *niwer* ‘my house’ in (618) is situationally unique given its link to the speaker. In (619), *to* marks a contextually unique referent. The utterance is made in a conversation about a certain manioc field that needs to be weeded. The referent of *es’hiwok* ‘weed’, mentioned for the first time, is therefore unique within the larger context of the manioc field.

(617) *teč* *worapik teč* *ses ro=aseroko-wo* [*to* *ses*].
 ART:REF.M come ART:REF.M sun 3SG.M=be.strong-COP ART:REF sun
 ‘The sun came up and it was getting strong.’
 Baure (Danielsen 2007: 448)

(618) *ro=kičo-wo=ro=hi* *ni=tori* *ni=ki’ino-wo* *pi=asko-ša* *ni=woyik-ša*
 3SG.M=say-COP=3SG.M=QUOT 1SG=friend 1SG=want-COP 2SG=help-IRR 1SG=make-IRR
 [*to* *ni=wer*]
 ART:REF 1SG=house?
 ‘He said: “My friend, would you help me make (repair) **my house**? ”’
 Baure (Danielsen 2007: 448)

(619) *ro=eporewan-čo-wo=ro* [*to* *es’hiwok*].
 3SG.M=cover-APPL-COP=3SG.M ART:REF grass
 ‘The weed is already covering it.’
 Baure (Danielsen 2007: 463)

¹⁰The present demonstrative *te/ti* is used with deictic and anaphoric referents, and can also introduce and establish new referents to the discourse, similarly to *this* in English. Danielsen (2007: 312) also notes that it often marks contrast. The use of the proximate demonstrative *teč/tič* is similar (Danielsen 2007: 313-314).

Examples (620) to (622) illustrate that *to* also occurs with referents that are specific, i.e. not unambiguously identifiable by the hearer.

(620) ro=ipkiek=ro ro=ina [to yakis].
 3SG.M=blow.down=3SG.M 3SG.M=use ART:REF stick
 ‘He blew him down with a stick.’
 Baure (Danielsen 2007: 429)

(621) teč ri=har-noki-wapa [to yiti].
 ART:REF 3SG.F=burn-mouth-COS ART:REF chili
 ‘She has burnt her mouth with chili.’
 Baure (Danielsen 2007: 461)

(622) ni=torak [to e-ser].
 1SG=find ART:REF GENPOSS-tooth
 ‘I found a tooth.’
 Baure (Danielsen 2007: 318)

The marker *to* also occurs with nonspecific referents which are no longer particular referents of their kind. The use of *to* in such contexts is illustrated in examples (623) and (624) below. The referent of *erapoe* ‘plantain’ in (623) and the referent of the first occurrence of *kahapo* ‘manioc’ are nonspecific because they occur within the scope of a question or negation. The second occurrence of *kahapo* ‘manioc’ in (623) is nonspecific because of the intensional predicate.

(623) pi=ahač=ri kwe-’i [to erapoe’].
 2SG=ask=3SG.F exist-EMPH ART:REF plantain
 ‘Ask her if there is plantain.’
 Baure (Danielsen 2007: 393)

(624) noka-wo [to kahapo] vi=hinoek-pa [to kahap].
 NEG-COP ART:REF manioc 1PL=search-DIR ART:REF manioc
 ‘There is no manioc, so we go and look for manioc.’
 Baure (Danielsen 2007: 199)

Hence, the preceding examples showed that *to* is compatible with all major referent types, occurring in definite, specific, and nonspecific functions. Furthermore, Danielsen (2007: 312) mentions that *to* occurs in a number of conventionalized constructions, e.g. in nominalizations. In (625) and (626), we see two examples that show the use of *to* marking a verbal expression as a referential expression.

(625) eto-a-ša-po [to pi=nik] pi=kač-poreiy-po.
 finish-LNK-IRR-PFV.REFL ART:REF 2SG=eat 2SG=go-REP-PFV.REFL
 ‘When you finish eating, do you go again?’
 Baure (Danielsen 2007: 464)

- (626) o=inoko-wo [to no=ačo-no-wo] to eton-anev tič
 3SG.M=resemble-COP ART:REF 3PL=have-NMLZ-COP ART:REF woman-PL:HUM ART:REF.F
 howe'.
 dolphin
 'The same what women have, has the dolphin woman.'
 Baure (Danielsen 2007: 458)

In addition, *to* is also formally integrated into the indefinite pronoun *to ka* (Danielsen 2007: 322), shown in (627) and (628).

- (627) a te kač moro'in=ro nka [to ka] ri=er.
 and ART:REF DIR be.thirsty=3SG.M NEG ART:REF INDEF 3SG.F=drink
 'And she got thirsty because there was nothing for her to drink.'
 Baure (Danielsen 2007: 456)

- (628) aiy ti koyepia-no ri=sokia [to ka] apo ri=koyepia-no.
 ay ART:REF converse-NMLZ 3SG.F=find ART:REF INDEF COMP 3SG.F=converse-NMLZ
 '(Ay) what a talkative person she is; she always finds something to talk about.'
 Baure (Danielsen 2007: 322)

All the previous examples illustrating the different uses of *to* showed that it is a versatile marker that marks a nominal expression as referential and occurs in different fixed constructions. The latter may suggest that *to* is not necessarily a referential marker but a syntactic nominal marker. Similarly to the referential article *te* in Rapa Nui, it can be argued that *to* marking an expression as a nominal one or as an argument follows from its referential function.

Nominal expressions used as predicates do not occur with *to*. This is shown in examples (629), (630), and (631). In contrast to what was shown for Rapa Nui, where the referential article occurred in opposition to a single nominal predicate marker, Baure uses a more complex system of different predicate markers that are suffixed to the nominal stem. Besides such predicate markers, nominal predicates also feature subject agreement markers (Danielsen 2007: 182). In other words, nominal predicates in Baure are used with verb-like morphology. Importantly, the article *to* does not occur in such contexts.

- (629) čas monči-wo=ni ni=kač ni=yon-poek šonoki-ye
 long.ago child-COP=1SG 1SG=go 1SG=walk-down way-LOC
 'In earlier times when I was a child, I walked barefoot that way.'
 Baure (Danielsen 2007: 428)

- (630) tin eton nka vi=moestar-wo=ri nka moestaro-wo=ri.
 ART:REF woman NEG 1PL=teacher-COP=3SG.F NEG teacher-COP=3SG.F
 'That woman isn't our teacher. She is no teacher.'
 Baure (Danielsen 2007: 195)

- (631) puhhh ver **howe-wapa=ri**.
 puuhhh PERF **dolphin-cos=3sg.f**
 ‘Puhhh, and she changed into a dolphin.’
 Baure (Danielsen 2007: 195)

7.2.3 Halkomelem

This Section deals with the article system of Musqueam Halkomelem, a Salish language spoken on Vancouver Island and the surrounding mainland along the Fraser River in both Canada and the United States. According to the Endangered Languages Project, there are only around 100 fluent speakers of Halkomelem left.¹¹ Halkomelem consists of several varieties which can be grouped into Island, Downriver, and Upriver dialects. The variety described in this section, Musqueam Halkomelem, belongs to the Downriver varieties.¹²

Musqueam Halkomelem (henceforth: Halkomelem) has a number of determiners that frequently occur with nouns. Although they are treated as separate articles with different deictic functions in Suttles (2004), I argue in this section that they can be analysed as exponents of a single referential article. An overview of these markers based on Suttles (2004: 340) is given in Table 7.3. Suttles (2004: 340) distinguishes between proximal (“present”) and visible, proximal (“nearby”) and nonvisible, and remote.¹³ The association of the article forms with the three deictic values of proximal & visible, proximal & invisible, and remote goes back to the demonstrative system, which has a set of demonstrative markers that can be grouped into these three categories (Suttles 2004: 351). The article exponents are grouped in the same way in Table 7.3 because most of the forms are weakened forms of the demonstratives. As will be shown in this section, the forms of the referential article may express some of those deictic functions in certain contexts, but they are by no means restricted to those functions.

Table 7.3: Halkomelem article exponents

	non-feminine	feminine
proximal & visible	tə, tʰə	θə
proximal & invisible	k ^w θə, k ^w ə, k ^w	ʔə, ʔ, k ^w ʔ(ə)
remote	k ^w ə, k ^w	k ^w sə

¹¹Cf. <http://endangeredlanguages.com/lang/2379>.

¹²For more details on the determiner system of Island Halkomelem which appears to be quite similar, see Gerdtts & Hukari (e.g. 2004). The determiner system in Upriver Halkomelem, on the other hand, shows greater differences to the Musqueam system discussed here; for details see Brown & Thompson (2013), Galloway (1993), Wiltschko (2002).

¹³(Suttles 2004: 348) lists another marker, λ, as an oblique article. This marker only occurs with proper nouns in oblique positions. Since its use is restricted to a subset of nouns and certain syntactic contexts, I do not regard it as a referential marker and hence do not treat it as an article here.

Examples (632) and (633) illustrate the deictic effects of the article. In (632a) the referent of *sq^wáméy̓* ‘dog’ is marked as proximal and invisible by using *k^wθə*. Suttles (2004: 342) notes that this sentence is only felicitous in a scenario in which the speaker is outside of the house but close to it, the dog being inside the house. Example (632b) with the article *tə* corresponds to a scenario where the speaker is outside the house, with the dog coming outside towards the speaker. In (633), the form *tə* (visible) is contrasted with the form *k^wə* (remote).

- (632) a. ni^ʔ sk^wtéx^w ʔə tə léləm [k^wθə sq^wáméy̓].
 be.there inside OBL ART:REF.VIS house ART:REF.NVIS dog
 ‘The dog is in the house.’
 b. ʔi ʔəmí ʔəλqəl [tə sq^wáméy̓].
 be.here come exit ART:REF.VIS dog
 ‘The dog came out.’
 Halkomelem (Suttles 2004: 342)

- (633) a. ni^ʔ ʔə [tə nə-ʔéθəqən]
 be.there OBL ART:REF.VIS my-front
 ‘in front of me (in a canoe)’ literally: ‘there at my front’
 b. ni^ʔ ʔə [k^wə nə-sliʔáʔaq^wt]
 be.there OBL ART:REF.REM my-rear
 ‘behind me (in a canoe)’ literally: ‘there at my rear’
 Halkomelem (Suttles 2004: 344)

Although the preceding examples showed that the forms of the article in Halkomelem are involved in marking deixis and visibility, their use is better accounted for by the analysis as different exponents of a single referential article. In fact, Gerdts & Hukari (2004: 159) draw a similar conclusion about articles in Halkomelem: “Semantically, articles refer to definite and indefinite NPs indiscriminately, as is general in Salish languages.”

Definite referents marked by the article are presented in examples (634) to (636). Example (634) shows the first mention of *smáyəθ* ‘deer’ in a story about deer hunting. In the context of deer hunting, the referent is contextually unique and thus definite. An anaphoric referent is shown in (635), which follows the sentence in (634) in the narrative. The referent of *ʔənsq^wáméy̓* ‘your dog’ in (636) is situationally unique. These three examples also show that both forms *k^wθə* and *tə* can be used to mark referents as definite.

- (634) λá ce^ʔ k^wáləx-t [k^wθə smáyəθ] wə-mí-əs cə^ʔ
 BE.3 FUT shoot-TR ART:REF.NVIS deer when-come-SUB:3 FUT
 wəl-cláq^w-θət.
 already-pass.through-self
 ‘They will be the ones who will shoot the deer when they start coming through.’
 Halkomelem (Suttles 2004: 525)

(635) ... wə-mí-əs wəlwél-ət-əs [k^wθə smáyəθ].
 ... when-come-SUB:3 be.chasing-TR-SUB:3 ART:REF.NVIS deer
 ‘[They’ll be the ones striking stones together] when they come chasing the deer.’
 Halkomelem (Suttles 2004: 525)

(636) ləq̄ ʔə q̄íq̄ək̄^wəls [tə ʔən-sq^wəméy]?
 habitually Q be.biting ART:REF.VIS your-dog
 ‘Does your dog bite?’
 Halkomelem (Suttles 2004: 471)

The following examples illustrate the use of the article with specific referents. Here, we see that the deixis parameter of the article exponents no longer straight-forwardly accounts for the use of the markers. The form *tə* (visible) is used to express the specific referent of *čičíʔq̄ən* ‘mink’ and of *k^wíʔx^w* ‘pitch’ in (637) and (638). Another example of a specific referent is provided in (639); the specific referent is marked by the form *k^wθə* (remote) of the article.

(637) s-ní^ʔ-ct wə-k̄^wəc-nəx^w [tə čičíʔq̄ən].
 NMLZ-AUX-OUR EST-see-TR ART:REF.VIS mink
 ‘Then we saw a mink.’
 Halkomelem (Suttles 2004: 347)

(638) k^wən-ət-əs [tə k^wíʔx^w].
 get-TR-3 ART:REF.VIS pitch
 ‘She took (some) pitch.’
 Halkomelem (Suttles 2004: 348)

The context of (639) is described by Suttles (2004: 348) as follows: “The conditional clause (d) [(639), LB] is from a statement made by Scouring Rush to Mink warning him of the dangers of standing beside him in the water. The log has not been previously mentioned in the narrative and so the article must be indefinite in English. The log is real though not yet visible to Scouring Rush and so takes the Halkomelem *k^wθə*; non-feminine near article.”

(639) ... wə-ʔəmí-əs [k^wθə θí q^wléy].
 ... COMP-come-SUB:3 ART:REF.NVIS big log
 ‘...when a big log comes.’
 Halkomelem (Suttles 2004: 348)

Suttles (2004: 345) notes that nonspecific referents are systematically encoded by the “remote” form of the article (*k^wə*). Examples of nonspecific referents are presented in (640) to (643). Note that the article surfaces as shortened *k^w* in (641). Example (643) shows that also nominalized predicates, here *nəsném* ‘my going’, can be referential and are marked by the article accordingly.

- (640) $nə-s-c-ǎíʔ$ $[k^{wə}$ $páy]$.
 my-NMLZ-do-value ART:REF.REM pie
 ‘I want some pie.’
 Halkomelem (Suttles 2004: 345)
- (641) $ʔə́y k^w$ $s-ném-ct$ $səwq-t$ $[k^w$ $pəwə́y]$ $ʔə$ $tə$ $cáwəcə́w$.
 good ART:REF.VIS NMLZ-go-our seek-TR ART:REF.REM flounder OBL ART:REF.VIS offing
 ‘We’d better go look for (any kind of) flounders offshore.’
 (Suttles 2004: 345)
- (642) $ʔéx^w-əs-t-Samx$ $čx^w$ $ʔə$ $[k^{wwə}$ $léctən]$.
 give-REC-TR-me you OBL ART:REF.REM knife
 ‘Give me a knife.’
 Halkomelem (Suttles 2004: 51)
- (643) $nə-s-cǎíʔ$ $[k^{wə}$ $nə-s-ném]$.
 my-NMLZ-do-value ART:REF.REM my-NMLZ-go
 ‘I want to go.’ (literally: ‘What I want is my hypothetical going.’)
 Halkomelem (Suttles 2004: 345)

From the examples given in Suttles (2004), it does indeed seem to be the case that nonspecific referents are systematically marked by $k^{wə}$, the remote form of the article. This suggests an alternative analysis of $k^{wə}$ as a separate nonspecific article. However, $k^{wə}$ is not restricted to occur with nonspecific referents. Its occurrence with definite referents was shown in (633b); more examples are given below. Example (644) shows that $k^{wə}$ can also mark an expression as specific. Here, $k^{wə}$ occurs with a verbal expression to mark it as referential. The referent of $k^{wə}$ $ʔi$ $yəsíqəs$ ‘something whistling’ is indefinite because the exact identity of the referent is not yet known by the speaker at the time of utterance. At the same time, the referent is part of an episodic past event, making it specific.

- (644) $s-niʔ-ct$ $wə-háyeʔ, ʔi$ ct $təw-x^wən-yə-ʔé.ý$
 NMLZ-AUX-our EST-leave AUX we somewhat-still-along-be.continuing
 $yə-ʔəxʔíʔxəl$ $ʔi$ ct $wəl-čléń-ət$ $[k^{wə}$ $ʔi$ $yə-síqəs]$.
 along-be.paddling.PL AUX we already-hear-TR ART:REF.REM AUX along-wistle
 ‘Then we left and we were still leisurely paddling along when we heard something whistling.’
 Halkomelem (Suttles 2004: 529)

Along with its function as remote marker, $k^{wə}$ has two other conventionalized functions that probably also go back to remoteness, distance, and uncertainty or irrelevance. For instance, $k^{wə}$ is commonly used to refer to referents in the past and deceased referents in particular (Suttles 2004: 343). Two examples of this use are given in (645) and (646).

(645) k^{wə} nəsílə
ART:REF.REM grandfather
 ‘my late grandfather / great uncle’
 Halkomelem (Suttles 2004: 343)

(646) wə-łá ʔal s-ńíwəyəl-s [k^{wə} nə-šx^wwéləy] ʔəy ni wə-mák^w ʔal
 EST-BE.3 just NMLZ-worship-POSS.3 **ART:REF.REM** my-parents and AUX EST-all just
 qáy.
 die
 ‘It was the only religion of my parents until they both died.’
 Halkomelem (Suttles 2004: 114)

Another use of the remote form of the referential article is to signal uncertainty or imprecision regarding the identity or exact properties of the referents. This is shown in (647) and (648).

(647) łél-x-əs tə tóməl ʔəy [tə stəʔé ʔə k^{wə} snás].
 apply-TR-3.TR **ART:REF.VIS** red.ochre and **ART:VIS** like OBL **ART:REF.REM** oil
 ‘He applied red ochre and something like oil.’
 Halkomelem (Suttles 2004: 70)

(648) stém [k^{wə} ləpláš] təʔí?
 what **ART:REF.REM** board this
 ‘What kind of board is this?’
 Halkomelem (Suttles 2004: 399)

The use of k^{wə} to signal that the exact properties of the referent are either not relevant as in (647) or unknown to the speaker as in (648) could also account for its use with nonspecific referents. Nonspecific referents often occur in contexts questioning the existence of a referent or in free choice contexts where the speaker signals that the exact identity of the referent is irrelevant (cf. example (640)). However, since k^{wə} also occurs with specific and definite referents, it is no nonspecific article. k^{wə} being the preferred or conventionalized form of the article used with nonspecific referents may be a consequence of its additional functions to express remoteness and uncertainty.

In contrast to its use with nominal referring expression, the referential article is not used with nominal predicates. Examples (649) to (652) illustrate this. In (649), (650), and (651), the nouns are used as predicates and are therefore not referring. In example (652), the expression *cí·tməx^w* ‘owl’ is used in a similitive construction and is non-referring as well.

(649) wəlθíʔə mə θəwłá.
ritualist CERT she
 ‘She is a ritualist.’
 Halkomelem (Suttles 2004: 382)

- (650) **spé'eθ** čə.
bear QUOT
 'It is said / supposed to be a bear.'
 Halkomelem (Suttles 2004: 374)
- (651) **nə-s-wá' ce' nə-sʔpíwən.**
 my-NMLZ-OWN FUT **my-shirt**
 'It will be my shirt.'
 Halkomelem (Suttles 2004: 375)
- (652) **stém stém yəx^w tθé' təw-cí·tməx^w wa'.**
 what what INFER DEM **like-great.horned.owl** PRESUMP
 'What can that be? It looks like an owl.'
 Halkomelem (Suttles 2004: 390)

In contrast to nominal predicates, the referential article does occur with nouns in equative constructions. This is shown in (653) and (654). Here, the expressions *nəhálk^w* 'pocket knife' and *spé'eθ* 'bear' are referring to a definite and a specific referent, respectively.

- (653) **ʔa mə [k^wθə nəhálk^w].**
 BE.3 CERT **ART:REF.NVIS** my-be.breaking
 'It is my pocketknife (that he has in his possession, as I noticed just now).'
- (654) **[k^wəw wə-spé'eθ] mə.**
ART:REF.REM EST-bear CERT
 'Of course it's a bear.' (which cause something such as a dog being mauled)
 or: 'What can you expect of a bear?' (as when one has broken into a cooler)
 Halkomelem (Suttles 2004: 384)

Halkomelem has flexible word classes, as is common for Salishan languages (cf. Beck 2013). This means that words which are typically referring can equally be used as predication and vice versa. Examples (655) to (658) show that typical predicates can be used as referring expressions when marked as such by the article.

- (655) **nəwə yəx^w ce' xá'ləm-ət [tə həyq^w] ...**
 be.you INFER FUT agree-TR **ART:REF.VIS** be.burning ...
 'You will (honour us to) be the one to look after the fire ...'
 Halkomelem (Suttles 2004: 376)
- (656) **stém k^wə [tə k^wənə-t-əx^w]?**
 what then **ART:REF.VIS** hold-TR-you
 'What is it you are holding?' literally: 'What is your holding?'
 Halkomelem (Suttles 2004: 385)

- (657) stém [kʷə ʔə-s-cʰiʔ]?
 what ART:REF.REM you-NMLZ-want
 ‘What do you want [e.g., of food laid out]?’
 Halkomelem (Suttles 2004: 400)
- (658) híθ [kʷə nə-s-ʔiməx].
 last.long ART:REF.REM my-NMLZ-be.walking
 ‘I was walking a long time.’ (literally: ‘My walking lasted long.’)
 Halkomelem (Suttles 2004: 420)

The preceding examples have shown that the three forms $tə$, $k^w\thetaə$, and $k^wə$ systematically occur with nominal expressions to mark them as referential. The reason for analysing them as forms of a referential article and not as e.g. demonstratives is that they do not only occur occasionally. On the contrary, a referential nominal expression is always marked by one of these forms or another determiner. Halkomelem has a complex system of additional demonstratives that can occur with nominal expressions to express (mostly) spatial deixis. Those demonstratives can also occur pronominally, while the three forms that I analyse as a referential article only occur adnominally (Suttles 2004: 350). From a syntactic point of view, a plausible alternative analysis is to treat $tə$, $k^w\thetaə$, and $k^wə$ as three distinct determiners within a larger system of determiners. I do not want to argue against such an analysis here. However, for the purposes of the present study, their referential functions determine whether they should be treated as forms of the same article or as separate articles. And based on their referential functions, $tə$, $k^w\thetaə$, and $k^wə$ cannot be analysed as separate articles. Although we saw that the three markers can express distinctions of deixis, remoteness, and certainty, this section also showed that a single form does not pattern with certain referential functions. Instead, the use of one of those three markers generally marked expressions as referential, without necessarily discriminating between single referential functions. Only the association of nonspecific referents with $k^wə$ appeared to be more systematic. I argued that $k^wə$ can nevertheless not be analysed as a nonspecific article because it also occurs with definite and specific referents. Therefore, $tə$, $k^w\thetaə$, and $k^wə$ can be analysed as different forms of a referential article whose main function it is to indicate that the expression they occur with is a referring expression.

7.3 Summary

This chapter presented two article types that cross-cut the definite and the indefinite domain: inclusive-specific and referential articles. Inclusive-specific articles were shown to be used to mark definite and specific referents without distinguishing between these. They do not occur with nonspecific referents. In systems with inclusive-specific articles, nonspecific referents can

either be expressed as bare nouns as was shown for Bemba. The alternative is to have another non-specific article that occurs with nonspecific referents, which was shown for Ayoreo and Tongan. Inclusive-specific articles are very rare crosslinguistically, occurring in 4 out of 104 languages in the sample, namely in Ayoreo, Tongan, Maori, and Siar Lak. This makes it difficult to make further generalizations, but it appears that such systems are strongly dispreferred in the absence of a nonspecific article. Out of these 4 languages and Bemba, which is not part of the sample, only Bemba does not make use of an additional article. More research on individual languages with inclusive-specific articles is needed to draw a more accurate picture about their stability.

Another finding of this chapter concerning inclusive-specific articles is the pattern found in Tongan, where the inclusive-specific article is typically replaced by a “stronger” anaphoric article in pragmatic definite contexts. If the use of the anaphoric article and the use of the inclusive-specific article pattern with pragmatic and semantic definite contexts as systematically as the data reviewed in this chapter suggests, then the Tongan article should be analysed as a weak inclusive-specific article. At this point, I am not aware of other languages with this pattern. Thus, more research is needed to assess the crosslinguistic status of weak inclusive-specific articles.

The other type of domain-crossing articles presented in this chapter were referential articles. Referential articles do not mark certain referential functions as such but signal that the nominal expression is a referential one; they occur with definite, specific and nonspecific referents. I presented three examples from Rapa Nui, Baure, and Halkomelem. What these three languages had in common is that the presence of the article was required for a nominal expression to be interpreted as referential. The referential articles in Rapa Nui and Baure consisted of a single exponent used in all referential contexts. The referential article in Halkomelem corresponded to a complex determiner system in the traditional sense. While the three relevant markers could be shown to have certain additional properties that the other ones in the system did not have, all three markers showed a great deal of functional overlap. Crucially, they were all shown to be able to encode various referential functions which made them indistinguishable on the basis of their referential functions. This motivated their analysis as different forms of a single referential article.

Chapter 8

Articles: Crosslinguistic trends and variation

In this chapter, I discuss a number of trends and patterns of variation of articles in the world's languages. Section 8.1 is concerned with the crosslinguistic distributions of different article types. Then, I show a number of areal trends in Section 8.2. A question related to the crosslinguistic distribution is the one of why generic articles are not attested. This question is taken up in Section 8.3. Section 8.4 then revisits referential scales based on the data discussed in the previous chapters. The last part of this chapter proposes a typology of articles (Section 8.5).

As was mentioned in Chapter 1, the sample contains 104 languages and 141 articles. The number of articles is higher than the number of languages, since some of the languages in the sample have more than one article. The number of 104 languages resulted from excluding several languages from a larger pre-sample of 150 languages. Languages were excluded because the descriptions did not contain sufficient examples of the (potential) article to be certain enough about its referential functions and properties.

The sample is controlled in a way in that languages from as many different families and geographical areas were included. Due to the unequal distribution of articles in the world's languages, the sample is nevertheless a convenience sample. At the same time, it aims at being a variety sample, trying to capture the variety of articles attested in the world's languages. As the sample only includes languages that have articles, more general conclusions about the distribution of articles as opposed to the absence of articles is not possible. All observations in this chapter are conditioned on the existence of articles in the language. Therefore, I will only discuss the observed numbers as such with no additional statistical testing. Statistical tests would only be conditioned on the availability of articles in a language in the first place, and would not offer much more insights in certain cases of rare article types, e.g. in the case of 4 inclusive-specific articles in the sample of 141 articles.

8.1 Overall distribution of the 10 article types

The frequency distributions of the 10 article types in the sample of 104 languages is presented in Figure 8.1. The article types distinguished are: definite (*def*), anaphoric (*ana*), weak definite (*def_{weak}*), recognitional (*recog*), exclusive-specific (*exspec*), nonspecific (*nspec*), indefinite (*indef*), inclusive-specific (*inspec*), weak inclusive-specific (*inspec_{weak}*), and referential (*ref*) articles. Since presentational articles are a subtype of indefinite articles in that they express the same referent types, they are included in the count of indefinite articles here. Note that the classification of the Tongan article (cf. Section 7.1.3) as a weak inclusive-specific is somewhat tentative. Table 8.1 gives an overview of the frequencies and proportions of each article type in the sample, sorted from most to least frequent.

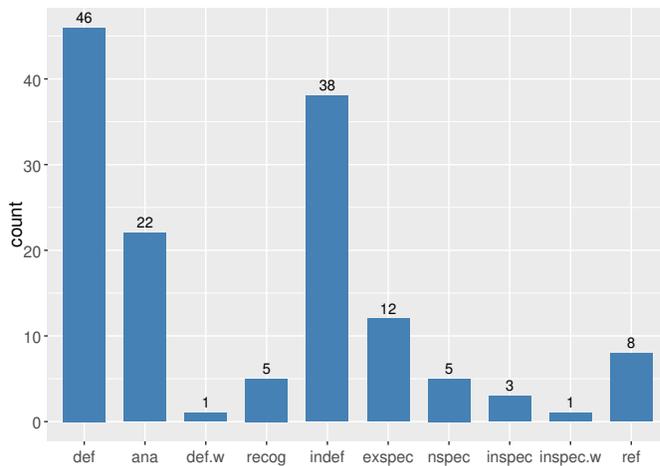


Figure 8.1: Distribution of the 10 article types in the sample

Table 8.1: Distribution of the 10 article types in the sample

article type	N in sample	proportion
definite	46	0.33
indefinite	38	0.27
anaphoric	22	0.16
exclusive-specific	12	0.09
referential	8	0.06
recognitional	5	0.04
nonspecific	5	0.04
inclusive-specific	3	0.03
weak definite	1	0.01
weak inclusive-specific	1	0.01

We see that definite articles, making up a third of the articles in the sample, correspond to the most frequent article type. This confirms an earlier observation by Dryer (1989, 2014) that definite articles are more frequent than indefinite articles. Comparing the frequency of those two types, however, shows that indefinite articles are not much less frequent than definite articles. Especially with the more fine-grained distinction of article types made in this study, the proportions of definite (0.33) and indefinite articles (0.27) are very similar if compared to the other types of articles. The third most frequent type is the anaphoric article, which makes up 16% of the sample. This makes anaphoric articles be one of the three most common types of articles in the world's languages. It confirms an earlier intuition of Dryer (2014), who at the same time draws a slightly different conclusion:

“My distinguishing anaphoric from nonanaphoric definites is motivated largely by the fact that articles coding anaphoric definites are very common among the world's languages, certainly more common than articles coding pragmatically or semantically nonspecific indefinites and probably more common than articles with a distribution close to that of either the definite or the indefinite article in English.” (Dryer 2014: 236)

The distributions of articles in the sample of the present study confirm that anaphoric articles are fairly common in the world's languages. This also makes them considerably more common than nonspecific articles, which correspond to “pragmatically or semantically nonspecific indefinites” in Dryer's terminology. Still, compared to the proportions of definite and indefinite articles in the sample, it seems that the latter two types are in fact more common than anaphoric articles, especially definite articles, whose number in the sample (46) doubles the number of anaphoric articles (22). The difference between Dryer's observation and the findings of the present study is very likely due to different definitions of articles, especially regarding anaphoric articles. I only included markers as anaphoric articles if the data presented in the descriptions (e.g. in sample text collections provided in grammars) suggested that the marker is used with anaphoric referents in a consistent way, i.e. if it was recurrent in such contexts (and only in such contexts). This excluded a large number of languages with (adnominal) anaphoric markers, where the data suggested that the markers can be used to signal an anaphoric referent but are not required to be used in many instances. This seems especially common in many Australian languages (cf. Section 8.2). Including such markers even though they may not be required to be used with anaphoric referents would indeed lead to a substantial increase in the number of adnominal anaphoric markers compared to definite or indefinite articles in the world's languages. The crosslinguistic distribution of anaphoric markers and articles as well as the factors that condition their use certainly deserve a closer look in future work both from a typological and a theoretical perspective.

Related to the crosslinguistic distribution of anaphoric (or strong definite) articles is the occurrence of weak definite articles, which occurs once in the sample of the present study, namely in

German (cf. Section 5.3). Weak definite articles appear to be extremely rare across languages, and the data discussed in this study confirm earlier observations that the presence of a weak definite article is dependent on the presence of an anaphoric (or strong definite) article in the language. This issue will be taken up in detail in Section 9.1.1. The dependency between weak definite and anaphoric articles is clearly not bidirectional; anaphoric articles are very frequent crosslinguistically and in no way dependent on the presence of another weak definite article in the system. This is in line with similar observations made by Jenks (2018) and Schwarz (2019).

The fourth type of articles in the definite domain is the recognitional article. It is also very rare crosslinguistically, with 5 occurrences in the sample. Recognitional articles are found in the Papuan languages Oksapmin and Lavukaleve, and in the Australian languages Bininj Kun-Wok, Gooniyandi, and Yankunytjatjara. Their distribution in the sample suggests that recognitional articles are an areal phenomenon. More research is required to examine whether this is the case or if their absence in other areas of the world reflects a bias in the language descriptions. What is certain is that the expression of recognitional referents has not received sufficient attention in both the typological and theoretical literature, Himmelmann (1996, 1997) and Diessel (1999) being notable exceptions. This is one of the reasons that recognitional articles were included as a type of articles in the present study; without their discussion in typological and theoretical work, it is very likely that they will go unnoticed in language documentation and grammar writing. To what extent dedicated recognitional markers other than the examples discussed here are used systematically or obligatorily with recognitional referents will also need to be examined in future research.

In the indefinite domain, indefinite articles are considerably more frequent than all other types, making up 27% of the sample. Out of the 38 indefinite articles, roughly half of them, 18, are what I called presentational articles, which are equivalent to other indefinite articles in their referential functions but whose use is restricted to discourse-prominent referents. This fairly equal occurrence of indefinite articles and presentational articles leads to two questions. Why are indefinite articles restricted to occur with discourse-prominent referents in some languages and why they are not in others? And why do we find this distinction with indefinite articles but not with definite articles?

Regarding the first question, the data from Lango (cf. Section 6.2.1) suggested that at least in some languages, the restriction to discourse-prominent referents can be a result of how information structure interacts with syntactic positions. Lango was shown to restrict the use of the presentational article to subjects, which are interpreted as topics and systematically occur in a clause-initial position. Given that the referents of subjects are typically also animate, discourse-old, and in control of the situation, their inherent and contextual properties make them generally highly discourse-prominent. Thus, when referring to a non-identifiable referent, we can expect

the potential of mistaking it as definite being high in subject positions but not in other syntactic contexts. It is therefore not surprising that the use of an indefinite article will not be extended to and conventionalized in other contexts, where the association with the identifiability of the referent is much weaker.

This is of course not restricted to languages like Lango, where the coding of information structure and referentiality is hard-wired into the syntactic structure. That identifiability is often correlated with topicality, jointly resulting in high discourse prominence, is well known. Building on that, one could argue that because discourse-prominent referents are generally expected to be identifiable and thus definite, an additional indefinite marker is required in those cases in which the identifiability status deviates from this default, ensuring that the referent is interpreted as non-identifiable. By contrast, less prominent discourse referents are less strongly associated with identifiability and may even be expected to be non-identifiable, so that their identifiability status does not require overt marking. Moreover, whether or not less prominent referents are identifiable by the hearer may simply be irrelevant.

If such an account is adopted, then the question becomes why some languages do extend the use of indefinite articles to referents that are not discourse-prominent. In other words, why do we not only find presentational articles but also indefinite articles? Addressing this question will require a more detailed quantitative assessment of the distributions of single indefinite and presentational articles, which goes beyond the scope of the present study. A factor that certainly plays a role is the paradigmatic relation between the indefinite/presentational article and other potential articles or determiners in the system. If a language has other articles or determiners, those may favor the extension of presentational articles to indefinite articles for syntactic rather than semantic reasons. However, based on the distributions in the sample, it is difficult to draw any conclusions in this regard. We find 12 languages with a single presentational article, while 6 languages have a presentational article and at least one other article. As for indefinite articles, 9 languages have a single indefinite article, and 10 languages have at least one other article. Clearly, more research is needed to disentangle linguistic and non-linguistic factors such as contact to account for the distribution of presentational vs. indefinite articles.

The second question regarding the restriction of articles to discourse-prominent referents concerned the observation that this does not seem to happen in the definite domain. This would fit in with the general picture of definite referents typically being discourse-prominent based on other inherent and contextual properties as well. However, we do find markers that are used with identifiable and thus definite referents in contexts that are determined by discourse-pragmatic factors. The only difference to the situation of presentational vs. indefinite articles is that I treat definite markers restricted to discourse-prominent referents as demonstratives, because this is a function that demonstratives across language generally have. Typical contexts are situations in which

an identifiable referent is re-introduced in the discourse, having been “inactive” for a number of utterances or being identifiable without having been mentioned at all in the current discourse situation. This is where (anaphoric) demonstratives but also recognitional articles come in. Thus, the speaker may use such markers exactly in those cases in which the referent is not very likely to be constructed as identifiable based on the immediately preceding discourse. As was mentioned for Hausa, Lakota, and Urama in Section 5.3, languages with anaphoric markers in addition to definite articles often use these anaphoric markers at the beginning of topic chains, in situations of topic shifts, and more generally for identifiable referents whose prior mention is further away or not part of the current discourse situation.

One may then ask why we do find so many definite articles that mark identifiable referents regardless of their discourse prominence. From a functional point of view, only less prominent referents may require explicit coding as identifiable. There might be a simple answer to this question. We know that there are at least as many (if not more) languages without articles than languages with articles in the world. In the WALS sample of articles with 620 languages in total (Dryer 2013b,c), Dryer classifies 216 languages as having a “definite word distinct from demonstrative” and 92 languages as having a “definite affix”. Together, these languages make up around 50% of the 620 languages, and given stricter criteria for what counts as an article, the proportions of languages with definite articles may be even lower. In other words, while definite articles correspond to the most frequent type of articles found across languages, articles in general are not very frequent and, returning to the question of discourse-prominent referents, many (if not most) languages do in fact use only demonstratives to mark less prominent definite referents as identifiable. As was argued for indefinite articles, the development of definite articles could also be influenced by additional, e.g. syntactic factors.

Returning to the other types of articles in the indefinite domain, exclusive-specific and nonspecific articles, the distributions in the sample suggest that these are less frequent across languages than functionally broader indefinite articles, with only 6 nonspecific and 12 exclusive-specific articles. There is no evident explanation for why some languages restrict an article in the indefinite domain to specific referents (exclusive-specific articles) or use it for both specific and nonspecific referents (indefinite articles). The latter type is crosslinguistically more frequent, and it may be the case that some of the exclusive-specific articles are “younger” articles that may develop into indefinite articles. This is not necessarily the case, however, and one should be careful in assuming that exclusive-specific articles are only an intermediate step towards the development of indefinite articles. The entire article or determiner system could play a role in their development, i.e. the use of other articles with referential expressions and the availability of bare nouns in argument positions. In the cases of Lakota and Biak, for instance, the exclusive-specific article is part of a system including a definite and a nonspecific article. This fits into the wider trend observed

in the sample; only 3 out of 12 languages (Angolar, Barwar, Rajbanshi) with an exclusive-specific article have no other article in their system. In the remaining languages, the exclusive-specific article is opposed to at least a definite or an anaphoric article.

Nonspecific articles are clearly very rare crosslinguistically, with 5 occurrences in the sample. Based on the languages in the sample and other cases that I am aware of, they require the availability of either an inclusive-specific or an exclusive-specific article in the system (cf. Section 9.1.2). Given such paradigmatic requirements, it is not surprising that nonspecific articles are very rare across languages.

Finally, the three types of domain crossing articles, inclusive-specific, weak inclusive-specific and referential articles are rare as well, with only 3, 1, and 8 occurrences out of 141 articles in the sample, respectively. Taken together, the four languages with inclusive-specific articles (Ayoreo, Maori, Siar Lak, Tongan) all have a nonspecific article as well. Bemba (cf. Section 7.1.1) is the only good counter-example that I am aware of.¹ The weak inclusive-specific article in Tongan may present further evidence for the instability of inclusive-specific articles. As was shown in Section 7.1.3, the data suggests that an additional marker is used together with the inclusive-specific article in pragmatic definite contexts, since the inclusive-specific article was no longer sufficient to mark the referent as identifiable. This resulted in the formation of what I analysed as an anaphoric article, and possibly led to the restriction of the former inclusive-specific article to semantic definite contexts, making it a weak inclusive-specific article. However, it seems likely that the Tongan system will develop into a new system with the anaphoric article as a new definite article and the weak inclusive-specific article as a new exclusive-specific article.

Referential articles, which make up 6% of the articles in the sample, are also very rare crosslinguistically. In some languages, they may be opposed to a predicate marker, both of which are part of a larger system of determiners. From the languages of the sample, this is the case only in Rapa Nui (cf. Section 7.2.1) and Sunwadia, both of which are Oceanic languages. The other languages in the sample allow bare nouns for non-referential nominal expressions (Basque, Halkomelem, Tephua), or use verbal morphology with the nominal expression in those cases (Baure, Mamaindé, Sabanê).² The early proposal of the development from article to noun marker in Greenberg (1978: 246-247) suggests that if languages have articles with very broad referential functions as is the case with referential articles, these articles may eventually become nominal markers, losing their

¹Given that the Bemba data presented is taken from Givón (1969), a study from over 50 years ago, and given Givón's focus on selected grammatical properties of Bemba, I cannot exclude the possibility that Bemba may have additional markers in the indefinite domain that are not described in Givón (1969) or in Hoch (1964), the only descriptive sources of Bemba that I am aware of. Therefore, I did not include the language in the sample. More research will be needed to determine to what extent the dependency between inclusive-specific and nonspecific articles is systematic.

²The use of "verbal" morphology also includes copulas and markers that are not necessarily used with typical verbs in the languages. Importantly, these are not markers that would form a system of determiners together with the referential article.

function as referential marker. While the present study showed that referential articles are recurrent in the world’s languages, future research is required to shed more light on their diachronic stability, which will help to explain why they are relatively rare crosslinguistically.

8.2 Areal trends

Figure 8.2 shows the distribution of article types across the six macro areas as used in WALS and Glottolog (Dryer & Haspelmath 2013, Hammarström et al. 2020): Africa, Australia, Eurasia, North America, Papunesia, and South America. Besides the overall lower number of articles in especially Australia and to a lesser extent in South America, we can observe five potential areal trends. Africa and North America show (i) a relatively high proportion of definite articles (shown in dark red). Eurasia and South America, on the other hand, feature (ii) a large number of indefinite articles (yellow). Related to that, we find (iii) the highest number of exclusive-specific articles (light green) in Africa. Australia stands out for (iv) its high proportion of anaphoric articles (red), and together with Papunesia it is the only area where (v) recognitional articles (light orange) are found. Regarding the last observation that recognitional articles appear to be restricted to Papunesia and Australia, Section 5.4.6 already mentioned potential explanations for their occurrence in certain types of speech communities. The other four areal trends will be discussed in turn in the remainder of this Section.

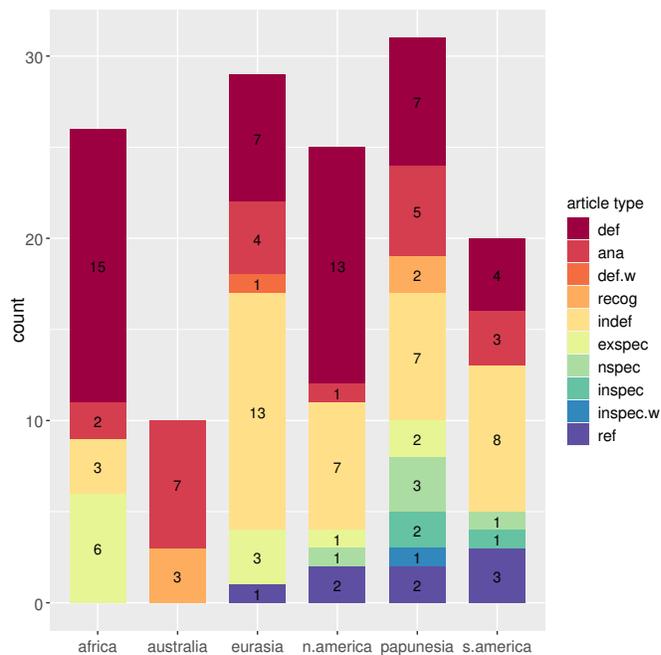


Figure 8.2: Distribution of the 10 article types across macro areas

Table 8.2 below summarizes the number of definite articles and their proportions in the sample across the six macro areas. These figures confirm a difference between Africa and North America and the other macro areas. For both Africa and America, definite articles correspond to over 50% of the articles in the sample, whereas they do not make up more than 25% in Eurasia, Papunesia, and South America. Australia is yet another case in that it does not feature any definite article. The absence of definite articles in Australia is part of a more general trend against articles in that macro area, which will be discussed in more detail below.

A explanation for the rather strong preference towards definite articles in Africa may be simply be a preference against articles in the indefinite domain, whose absence automatically could lead to higher proportions of definite articles. Given that definite articles are the most frequent type of articles in general, this trend may not be something that needs to be accounted for as such, and it may simply represent the default situation. This fits in with the distributions in Eurasia and South America, which have comparatively high proportions of indefinite articles. Papunesia does not necessarily have many indefinite articles but it is the area with the highest number of less frequent article types such as referential, inclusive-specific, nonspecific, and recognitional articles. This also contributes to the overall lower proportion of definite articles in that macro area.

Table 8.2: Distribution of definite articles across macro areas

macro area	N ART:DEF	N ART	proportion
Africa	15	26	0.58
North America	13	25	0.52
Eurasia	7	29	0.24
Papunesia	7	31	0.23
South America	4	20	0.20
Australia	0	10	0

The second trend that emerges from the distribution presented in Figure 8.2 is the high proportion of indefinite articles in Eurasia and South America. The distribution of indefinite articles across macro areas is summarised in Table 8.3.

Table 8.3: Distribution of indefinite articles across macro areas

macro area	N ART:INDEF	N ART	proportion
Eurasia	13	29	0.45
South America	8	20	0.40
North America	7	25	0.28
Papunesia	7	31	0.23
Africa	3	26	0.12
Australia	0	10	0

Table 8.3 confirms that the proportions of 0.45 and 0.40 of indefinite articles in Eurasia and South America are substantially higher than the ones of Papunesia (0.20) and Africa (0.12). Especially for Eurasia, these numbers suggest that indefinite articles are an areal effect. This could either be due to similar other grammatical properties that somehow favor the development of indefinite articles, or it could be due to contact and diffusion, or a combination of both. While a more detailed account would go beyond the scope of the present study, contact is very likely to play a role. Articles in Europe are known to be an areal phenomenon (cf. Leiss 2000, Schroeder 2006). Many of the languages with indefinite articles in Eurasia belong to Indo-Aryan, Turkic, and Mongolic, or are spoken in areas in which other Indo-Aryan languages are spoken as well. If not direct contact, diffusion over a longer period of time is likely to have played a role.

Contact and diffusion may also account for the relatively large number of exclusive-specific articles in Africa. Out of the 12 exclusive-specific articles in the sample, 6 are found in Africa, 3 in Eurasia, 2 in Papunesia, and 1 in North America. Two of these, the one of Lakota (North America) and Biak (Papunesia) are part of a more complex system with nonspecific articles. Thus, out of 10 remaining exclusive-specific articles with no nonspecific counterpart, 6 are found in Africa, 5 of which in West African languages spoken in Ghana, Nigeria, Cameroon, Angola. Of course, more detailed work focusing on this area is required to be certain, but the observations made so far are suggestive that areal effects may play an important role for the availability of articles.

The fourth trend that emerged in Figure 8.2 concerned the presence of anaphoric articles being preferred in Australia compared to the other macro areas. The distribution of anaphoric articles across the six macro areas is summarised in Table 8.4. Indeed, anaphoric articles correspond to 73% of all articles in the sample from Australia, while making up 15% or less of the articles in all other macro areas.

Table 8.4: Distribution of anaphoric articles across macro areas

macro area	N ART:ANA	N ART	proportion
Australia	7	10	0.70
South America	3	20	0.15
Eurasia	4	29	0.14
Papunesia	5	31	0.16
Africa	2	26	0.08
North America	1	25	0.04

A partial explanation of the preference towards anaphoric articles in Australia involves personal and demonstrative pronoun systems. 5 out of the 7 Australian languages with anaphoric articles in the sample either lack distinct third person pronouns (and use demonstratives) or the pronouns are described as being rarely used in speech or restricted in some way. This is summarized in

Table 8.5. A similar distribution can be seen in the WALS chapter on third person pronouns and demonstratives (Bhat 2013), with languages in Australia having a comparatively high number of formally related third person pronouns and demonstratives. Also Bhat (2004: 103) notes that “most of the Australian languages are two-person languages in which [sic] third person pronoun is primarily a demonstrative.”

Table 8.5: Third person pronouns in Australian languages with anaphoric articles

language	properties of third person pronouns	
Mangarrayi	no distinct third person pronoun	(Merlan 1989: 52)
Jingulu	pronoun forms only available in ACC and GEN	(Pensalfini 1997: 242)
Guragone	pronoun is rarely used	(R. Green 1995: 130)
Bardi	pronoun is rare in speech, no anaphoric use	(Bower 2004: 42)
Martuthunira	use of the pronoun is restricted	(Dench 1994a: 100)
Wubuy	no reported restrictions	(Heath 1984: 241)
Yuwaalaraay	no reported restrictions	(Giacon 2014)

Obviously, the lack of a third person pronoun does not directly lead to the presence of anaphoric articles. However, the development of an anaphoric article (which always go back to an anaphoric pronominal form)³ could be favored by the absence of a distinct third person pronoun and by the availability of complex demonstrative systems with dedicated pronominal anaphoric forms available in the system as well. The latter would make the use of demonstratives and the anaphoric marker more frequent, which in turn could favor the extension of an anaphoric pronominal marker to an adnominal marker. Another interesting property of Australian languages was observed by Bhat (2004: 191):

“They [Australian languages] do not show any distinction between nouns and adjectives and, correspondingly, they use the same set of proforms either as adnominals or as pronominals.”
(Bhat 2004: 191)

This more general property of not distinguishing between nominal and adnominal uses may have facilitated the extension of a pronominal anaphoric marker to an anaphoric article, and it could be another factor that has led to the high number of anaphoric markers in Australian languages.

It is obvious from the distributions articles in the sample that the high proportion of anaphoric articles in Australia is only one side of the coin. Not only are anaphoric articles very frequent in

³All anaphoric articles in the sample can also occur as pronominal forms. While I am not aware of any languages with an anaphoric article (or an adnominal anaphoric marker, for that matter) but no pronominal anaphoric marker, the opposite is commonly found in different areas of the world, e.g. in Chukchi (Dunn 1999: 96), Gamilaraay (Giacon 2014: 192), and Bantawa (Doornenbal 2009: 94).

Australia compared to other areas, but Australia generally shows a strong bias against articles. This is also the reason for the low number of Australian languages in the sample, being in line with earlier observations. For instance, Dixon (2002: 66) notes that “Australian languages do not have any class of ‘articles’, definiteness generally being shown by demonstratives or inferable from discourse structure”. Louagie (2019) implicitly comes to a similar conclusion. She offers a large-scale comparison of the structure of noun phrases in Australian languages with a sample of 100 languages. Out of those, she notes that two languages have markers called articles in the grammars (Louagie 2019: 184). However, in neither case do these markers correspond to articles in the sense of the present study. The first so-called article is a topic marker in Marra (Mangarrayi-Maran). The second one is a marker in Mawng (Iwaidjan) that has different functions which rather seem to be related to information structure than marking referentiality (Forrester 2015).

Louagie (2019: 184) further notes that a small number of languages can use demonstratives or personal pronouns adnominally with similar functions to definite articles in other languages. Two examples are the adnominal definiteness marker in Worrorra (Worrorran), which goes back to a demonstrative (Clendon 2014: 421), and a third person pronoun in Nyulynul (Nyulnyulan), which can be used adnominally to express that the referent is definite. According to the examples provided in Clendon (2014), the Worrorra marker appears to be used mostly as an anaphoric marker that can also mark referents as situationally unique (and potentially recognitional). It does not occur very systematically with nouns and is rather used as a pronoun in the text collections provided (Clendon 2014: 467-474). Therefore, I would not analyse it as an article. The marker *kinyingk* in Nyulnyul (McGregor 2011: 126) appears to be similar to an anaphoric article as well, but its use is not required to mark an anaphoric referent and rather resembles the use of the additional anaphoric markers in Hausa and Lakota, re-activating referents and marking topic changes. McGregor (2011) describes its use as follows:

“However, as distinct from the English word, it does not contrast with an indefinite article, and its use is not obligatory on definite or specific NPs: *yiil* ‘dog’ can be used in reference to a definite animal, and can translate as ‘the dog’; it can also translate as ‘a dog’. Examination of the contexts of use of *kinyingk* DEF in the textual corpus reveals that it is generally textually salient entities that are referred to by NPs with this determiner. Often these entities are the main protagonists in narratives, and the main topics of expository texts. But sometimes they are of more local importance, as for instance in line (34) of Text 3: from this point on the water in which the sandalwood has been steeped takes on a significant role. It is presumably for this reason that *kinying* DEF has a stronger identifying potential than *the*, and usually translates as ‘this’ or ‘that’.”

One could argue that those restrictions are similar to the restrictions of presentational articles in the indefinite domain, and that such anaphoric markers should also be treated as articles. I

do not analyse them as articles because such uses correspond to uses of demonstratives across languages in general. A quantitative corpus study examining the distribution of anaphoric articles and related anaphoric markers would go beyond the scope of the present study. In general, this may be required to properly assess to what extent anaphoric articles that appear systematically with anaphoric referents differ from other anaphoric markers or demonstratives whose use is more restricted by discourse-pragmatic factors.

A potential reason why Australian languages do not seem to use anaphoric or other definite markers in a more “systematic” way is that these markers are not used primarily to encode referential functions but discourse-structural ones. For instance, topicality being highly correlated with definiteness, it is not surprising that a marker that signals topicality is at the same time often used with referents that are definite.

Although noting that this is not necessarily common in Australian languages, B. Baker (2008) discusses topical markers that he calls articles in Wubuy and Ngakalan (Gunwinyguan), and Marra (Mangarrayi-Maran). He concludes that these markers “appear to have functions which are associated with discourse status interpretations [...]” (B. Baker 2008: 163). He shows that the markers occur in certain polarity contexts on the one hand and with anaphoric referents on the other, which is certainly not common. However, their use is similar to the distribution of *kinying* in Nyulnyul and other anaphoric markers in many Australian languages, whose use appears to be strongly guided by discourse-pragmatic factors. Their referential functions, on the other hand, only seem to be secondary.

Another characteristic of Australian languages that has been related to information and discourse structure is that of flexible noun phrase structure. Australian languages have long been cited for lacking noun phrase constituency (e.g. Blake 1983, Heath 1986, Nordlinger 2014, and references therein). More recent work, on the other hand, has revealed that this is not entirely so; Louagie & Verstraete (2016) showed in a crosslinguistic study that there is clear evidence for constituency in the noun phrases of many Australian languages. Still, many Australian languages additionally allow for flexible orders of the elements in the noun phrase and for discontinuous noun phrases. Especially the latter property was argued to be conditioned by discourse-pragmatic factors in e.g. Schultze-Berndt & Simard (2012) for Jaminjung (Mirndi). This suggests that the coding of information-structural and related functions generally plays an important role in many Australian languages. Due to the large functional overlap with coding referential functions, there may not be much functional pressure to encode referentiality in addition, which would account for the absence of articles as referential markers in many Australian languages.

Another related issue is the association between languages with a lower degree of syntactic constituency of nominal expressions and the availability of articles, mentioned by Himmelfmann (1997: 156). He observes that languages with “lose” nominal syntagmas generally do not feature

highly grammaticalized determiners like definite articles.⁴ This also relates to the absence of articles in Australia. Having proposed a number of potential factors that play a role in Australia's general lack of articles, more research is needed to properly disentangle those factors and to examine the role of language contact and diffusion.

8.3 The absence of generic articles

It is well known that different types of articles can be used in generic expressions (Carlson & Pelletier 1995, G. N. Carlson 2005, Dryer 2014, Givón 1984, Krifka et al. 1995). Often, generic subjects can be expressed using a definite article with a singular noun, or bare singular or bare plural nouns. This section will give a brief overview of the use of articles with generic referents, largely confirming that different types of articles may be used with generic expressions, and that there is no crosslinguistic evidence for the existence of generic articles.

The notion of genericity is usually used in two different senses: as reference to kinds (as opposed to reference to objects) and as the denotation of a general property (Krifka et al. 1995: 2). While not always analysed as referring expressions, there are proposals to treat generic expressions as referring. In this vein, Givón (1984: 440) notes:

“The facts of natural language, as we shall see below, tend to suggest that generic subject expressions [...] are in fact bona fide referring expressions, but referring to a group or a type rather than to an individual.” (Givón 1984: 440)

This is very similar to the description of generic referents in Chierchia (1998a: 351): “So a kind can be manufactured out of a property by taking the largest member of its extension (at any given world).” Generic expressions in the sense of reference to kinds are shown in (659).

- (659) a. **The potato** was cultivated in South America.
b. **Potatoes** were introduced into Ireland by the end of the 17th century.
c. The Irish economy became dependent upon **the potato**.
(Krifka et al. 1995: 2)

Example (660) shows generic statements that express a general property or a characterization of some sort. These are generalizations that involve the entire predication and not only the nominal

⁴Himmelman (1997: 156) puts it as follows: “Des weiteren ist mit der Hypothese die empirische ‘Vorhersage’ impliziert, daß es in Sprachen, in denen nominale Syntagmen generell ‘locker’ strukturiert sind (wie z.B. im Lateinischen und Nunggubuyu), keine stark adnominal grammatikalisierten D-Elemente (wie z.B. Definitartikel) gibt. Meines Wissens ist auch diese Behauptung empirisch korrekt. [Furthermore, this hypothesis implies the empirical ‘prediction’, that languages, where nominal syntagmas are structured in a generally ‘lose’ way (e.g. in Latin and Nunggubuyu), have no strongly grammaticalized adnominal D-elements (e.g. definite articles). According to my knowledge, this assumption is empirically correct as well.]” (the translation is mine).

expression in question; such characterizing sentences are also called gnomic, general, or habitual expressions (Krifka et al. 1995: 3).

- (660) a. John smokes a **cigar** after dinner.
 b. A **potato** contains vitamin C, amino acids, protein, and thiamine.
 (Krifka et al. 1995: 3)

As can be seen from (659) and (660), kind referring generic expressions are those in which definite articles or plural nouns occur, whereas some languages can use indefinite articles in characterizing expressions. Articles whose primary function it is to mark generic referents do not seem to be attested in the world's languages. While we find that generic referents are typically expressed by bare nouns in the plural or nouns in the singular together with a definite article in many European languages, many languages do not use articles in such contexts.

Across languages, there is no strong tendency with regard to the use of certain article types with generic referents. In general, we find definite, indefinite, inclusive specific, and referential articles that can be used to express generic referents. Other article types do not seem to be attested in such contexts. Two examples of definite articles that are used with generic referents come from Armenian and Bullom So. According to the description in Dum-Tragut (2009), examples (661) to (662) show two contexts with generic referents in Armenian (Indo-European, Armenia). We see that the definite article is used together with a singular noun. Example (663) shows that Bullom So (Atlantic-Congo, Sierra Leone, Guinea) can also express generic referents as a plural noun together with the definite article.

- (661) bžišk-ě piti ušadir lin-i.
 doctor-ART:DEF part careful be-DEB.FUT.3SG
 'A doctor must be careful.'
 Armenian (Dum-Tragut 2009: 111)

- (662) mard-u-n kar-oł ē krakel-ov č'-span-es bayc
 person-DAT-ART:DEF can-PTCP:PRS is shoot-INF-INSTR NEG-kill-FUT.2SG but
 xosk'-ov span-es.
 word-INSTR kill-FUT.2SG
 'One cannot kill a human with shootings, but with words you kill him.'
 Armenian (Dum-Tragut 2009: 162)

- (663) [imúsmènè àcé] nà cé-én àwòyén.
 water.cat ART:DEF 3PL COP-NEG frightening
 'Water cats are not scary.'
 Bullom So (Childs 2011: 72)

Also inclusive specific articles can be used to express generic referents. Example (664) shows that Bemba (cf. Section 7.1.1) can express a generic referent with the inclusive-specific article and a plural noun.

- (664) i-m-bwa nii-nama.
 ART:INSPEC-CL9-dog CL9-animal
 ‘Dogs are animals.’
 Bemba (Givón 1969: 47)

Basque (Isolate, Spain, France) and Baure (cf. Section 7.2.2) have referential articles. Examples (665) and (666) show the use of the referential article *-a* with generic referents in the singular and plural in Basque. The use of the referential article *to* in Baure is shown in (667).

- (665) haritz-a zuhaitz saindu-a da.
 oak-ART:INSPEC tree sacred-ART:INSPEC is
 ‘The oak is a sacred tree.’
 Basque (Hualde & Ortiz de Urbina 2003: 120)

- (666) txakurr-a-k ugaztun-a-k dira.
 dog-ART:INSPEC-PL mammal-ART:INSPEC-PL are
 ‘Dogs are mammals.’
 Basque (Hualde & Ortiz de Urbina 2003: 119)

- (667) ni=ipiko-wo [to tiporek].
 1SG=be.afraid-COP ART:REF chicken
 ‘I am afraid of chicken.’
 Baure (Danielsen 2007: 130)

In other languages, the use of articles is not compatible with generic contexts. For instance, both Sheko (Dizoid, Ethiopia) and Oko (Atlantic-Congo, Nigeria) have definite articles. Their definite articles are shown in (668) and (669), respectively.

- (668) só òtì hás-tà kì-b-tà ñ=sā-fín-á-mə há=ge-ñ
 up.there cow PROX.M-LOC exist-REL-LOC 1SG=arrive-descend-put-IRR-STI 3MS=say-DS
 òyt-ñ datà íf=sàk-ñ.
 cow.F-ART:DEF near.LOC 3PL=arrive-DS
 ‘He said: “Up there where there is a cow I will arrive and descend. They arrived near the cow ...”’
 Sheko (Hellenthal 2010: 144)

(669) *íkén óbèn wàmọ nẹnẹ è-gbe ka [úúúò àyẹ] á-wà érunrò*
 town ART:EXSPEC.SG exist REL S:3SG-BE COMP sheep ART:DEF.SG S:3SG-COP farmer
nẹnẹ fọrẹ na.
 REL surpass RC
 ‘There was this town in which Sheep was the greatest farmer.’
 Oko (Atoyebi 2010: 278-279)

Nouns that express generic referents, however, cannot be used with the definite article and occur as bare nouns. This is shown in (670) for Sheko and in (671) for Oko.

(670) *gárgá íntfù-rà há=gyá-mə.*
 termite wood-ACC 3SG.M=chew-IRR
 ‘A termite eats wood.’
 Sheko (Hellenthal 2010: 142)

(671) *ógbén e-yíwo.*
 child PROG-cry
 ‘A child is crying / Children are crying’
 Oko (Atoyebi 2010: 161)

Section 6.1.3 presented the indefinite article *-gə* in Bonan. As is shown in (672) and (673), the indefinite article is not used with generic referents, which are expressed as bare nouns in the singular. Similarly, Palula has an exclusive-specific article *áa/ák* (cf. Section 6.3.3). As can be seen in example (674), generic referents are expressed by bare nouns in Palula as well.

(672) *ɲantçhə=da lawa zaija bi-saŋ.*
 past=LOC shaman strong COP.S-EPIST
 ‘In the past, shamans were very powerful.’
 Bonan (Fried 2010: 83)

(673) *talo naŋda nɔvi doɣkəsa hkamo wa.*
 tall.building in dog raise difficult COP.O
 ‘Keeping dogs in an (apartment) building is difficult.’
 Bonan (Fried 2010: 83)

(674) *amzarái muɾ-u=bhaáu insaán na kha-áan-u*
 lion die.PTCP-M.SG=ADJ human.being NEG eat-PRS-MS.G
 ‘A lion doesn’t eat a human being that is dead.’
 Palula (Liljegren 2016: 398)

These examples show that the use of articles in generic expressions fundamentally differs from their uses in other referential contexts. Because the occurrence of articles in generic expressions does not appear to be restricted to a certain type of articles, we can assume that is not their primary

referential function that makes them compatible with generic contexts. That it is not the article that marks the referent as generic is supported by the fact that single languages do not necessarily have one consistent strategy to mark generic referents but usually use different strategies. We can thus formulate the following Universal.

Universal 1 *The absence of generic articles*

There are no articles whose main referential function is the coding of generic referents.

Part of the explanation for the absence of generic articles may be that generic contexts do not form a homogeneous group; as was mentioned above, we can distinguish between the two main types of kind-referring and characterizing generic contexts (and a more fine-grained distinction of more subtypes is possible). Especially the latter type is not interpreted as generic based on the nominal expression but on the entire predication, since it does not describe single events but general facts, statements, etc. (Krifka et al. 1995: 12). Traditionally, certain predicates such as ‘be extinct’, ‘be invented’, or ‘be common’ have also been treated as kind-selecting predicates (Chierchia 1998b: 363) because they evoke a generic interpretation. Therefore, at least characterizing generic expressions can plausibly be excluded as the type of contexts in which a nominal marker for genericity can develop.

This does not explain why generic articles do not develop in kind-referring expressions. One reason may be that there simply are no “suitable” source elements from which a generic article could develop. The most common sources of articles, demonstratives, pronouns, and the numeral ‘one’ are all used for individuation. In other words, emerging articles both in the definite and indefinite domain start out from markers that are used to draw the hearer’s attention to the referent, to the identifiability of an instantiated referent or simply to its discourse prominence. Kind-referring generic expressions appear to be exactly the opposite in that they do not make reference to instantiated referents but to the abstract concept or the kind that the nominal element expresses.

The development of a generic article would therefore require an article or other type of determiner that already occurs with nouns and is compatible with generic expressions like the articles mentioned in the beginning of this section. If another article were to develop in that language, replacing the first article in all contexts except generic one, a system could arise in which one of the articles is strongly associated with generic contexts so that it becomes a generic article. While we cannot say that such a development is impossible, it does not seem to be attested. This is not to say that there are no grammatical means to express genericity, but they are rather found in the verbal domain. For instance, many languages have dedicated habitual or gnomic present markers which are especially used with generic expressions that are not referring to single events or that depict general facts.

8.4 Referential scales: Evidence from articles

8.4.1 A refined referential scale

Returning to the referential hierarchies or scales presented in Section 3.4.2, we can now combine the functional domains of all article types discussed in Chapters 5, 6, and 7 to refine the referential scale and reassess the structure between the various referent types on the basis of the distribution of articles in the world's languages. Table 8.6 shows the referential functions of the 9 different types of articles. The relevant referential functions are: deictic (R_{dei}), anaphoric (R_{ana}), recognitional (R_{recog}), establishing (R_{establ}), relational bridging ($R_{\text{rel-bridge}}$), situationally unique ($R_{\text{sit.uni}}$), contextually unique ($R_{\text{cont.uni}}$), unique bridging ($R_{\text{u-bridge}}$), specific (R_{spec}), and nonspecific (R_{nspec}). We can additionally distinguish non-referential contexts (NONREF) in opposition to all other referential contexts, similarly to the distinction of non-argumental expressions in the referentiality scale proposed by von Heusinger (2008). What I mean by non-referential contexts largely corresponds to nominal expressions used as predicates. The order of the referential contexts in Table 8.6 is motivated by how they are encoded by different articles and other referential markers. The overall scale from pragmatic definite, semantic definite, specific, nonspecific, and non-referential contexts is not new and in line with previous accounts. What is new is the integration of a more fine-grained distinction in the definite domain.

Deictic referents are located at the pragmatic definite end of the scale. This is warranted by how demonstratives (DEM) encode deictic referents and how different article types in the definite domain are compatible with their expression. Deictic uses constitute the main or prototypical contexts of demonstratives, and some demonstratives are used to mark deictic referents exclusively. Therefore, all other referential contexts are marked as optional for demonstratives in Table 8.6.

Table 8.6: The distribution of (referential) markers across the referential scale

marker	R _{dei}	R _{anaph}	R _{recog}	R _{establ}	R _{rel-bridge}	R _{sit.uni}	R _{cont.uni}	R _{u-bridge}	R _{spec}	R _{nspec}	NONREF
DEM	✓	(✓)	(✓)	(✓)	(✓)						
ANA		✓	✓	✓	(✓)	(✓)					
RECOG			✓	(✓)							
DEF _{WEAK}						✓	✓	✓			
DEF		✓	✓	✓	✓	✓	✓	✓			
INSPEC		✓	✓	✓	✓	✓	✓	✓	✓		
REF		✓	✓	✓	✓	✓	✓	✓	✓	✓	
EXSPEC									✓		
NSPEC										✓	(✓)
PRED											✓

Although they cover the definite domain, I argued that definite (DEF), inclusive-specific (INSPEC), and referential (REF) do not mark referents as deictic as such. They may rather be compatible with deictic referents, but mark them as situationally unique. Therefore, their functional domains do not include deictic uses in Table 8.6. Anaphoric (ANA) articles are not compatible with deictic referents. Because anaphoric referents can also be encoded by demonstratives, however, anaphoric uses are situated next to deictic ones on the scale. As was shown in Section 5.2, anaphoric articles usually also encode recognitional and establishing referents. This is why these two functions are located next to anaphoric uses on the scale. The fact that we find recognitional articles which can also be used to (re)establish discourse referents (cf. Section 5.4.6) is further evidence that recognitional and establishing uses are similar and should be in adjacent positions on the scale. Returning to anaphoric articles, we saw that the anaphoric article in Akan (Section 5.2.3) can also be used to mark rel-bridging and situationally unique referents.⁵ Therefore, those two uses are the next ones on the scale. The order of rel-bridging and situationally unique uses is motivated by the distribution of weak definite (DEF_{WEAK}) articles. Weak definite articles can mark referents as situationally unique but not as rel-bridging. Regarding the traditional distinction between pragmatic and semantic definites, situationally unique referents appear to be between those two types of definites; or, what is probably more likely, situationally unique referents may be both at the same time, and languages or speakers may construct them as either (or both) in a given instance. On the one hand, situationally unique referents rely on the context of the discourse situation,

⁵This may also hold for the anaphoric article in Hausa (Section 5.3.2), which at least also seem to be used with situationally unique referents. More detailed descriptions of especially rel-bridging contexts are necessary to draw a clearer picture.

which is a property of pragmatic definites. On the other hand, those referents are constructed as unique (within the immediate discourse situation), and uniqueness is a property associated with semantic definites.

The observation that situational uniqueness is in an intermediate position between pragmatic and semantic definites relates to the proposal by [Himmelman \(1997\)](#), who argues that recognitional and establishing rather than anaphoric contexts are the relevant contexts for the development of definite articles. The results of the present study confirm this; moreover, they suggest that situational uniqueness is the next relevant context that articles in the definite domain can extend their functions to.

Contextually unique and u-bridging referents have to be located at the semantic end of definite functions, given that only definite, weak definite, inclusive-specific, and referential articles encode them, as opposed to demonstratives and anaphoric articles. The order of contextually unique and u-bridging referents in [Table 8.6](#) is arbitrary, since the use of articles does not appear to distinguish between them. This confirms the proposal of [Schwarz \(2009\)](#) to treat u-bridging referents as a type of contextual uniqueness.

The specific use, the next function on the scale, is adjacent to the semantic definite functions because of inclusive-specific and referential articles which are able to encode both types of functions. The next and last referential function on the scale are nonspecific uses. Their position is motivated by the functions of indefinite articles, used for both specific and nonspecific contexts. The main order between definite, specific, and nonspecific is not new and confirms a distinction that is part of all previous referential scales or hierarchies.

Following nonspecific uses, we can locate non-referential uses at the very end of the referential scale. This is motivated by the distribution of articles in two ways. We find referential articles (cf. [Section 7.2](#)), which cover the entire referential space (excluding deictic referents) but which are not used with non-referential nominal expressions. Interestingly, we also find nonspecific articles that may be used with nominal predicates (also in contexts which may be ambiguous between a nonspecific referential or irrealis predicate interpretation). Such a marker is found in Maori and will be described in [Section 9.3.3](#). The referential scale that falls out from [Table 8.6](#) is shown in [\(675\)](#), from most referential (or definite) on the left to non-referential on the right.

(675) **Referential scale**

deictic > anaphoric > recognitional, establishing, rel-bridging >
situationally unique > contextually unique, u-bridging >
specific > nonspecific > non-referential

The refined referential scale essentially corresponds to the previous referential hierarchies or scales. In the definite domain, it combines insights from earlier work, distinguishing between

pragmatic and semantic definites, and integrating bridging referents. Importantly, the use of articles showed that situationally unique referents should not be grouped together with deictic referents, even though their contexts may often overlap. Situationally unique referents are expressed by weak definite articles, but they can also be marked by anaphoric articles in some languages. The second important innovation is the integration of recognitional and establishing uses into the larger picture of the referential scale. Another important result from the use of articles is that bridging contexts do indeed not form a homogeneous group but need to be divided further. Rel-bridging is largely similar to anaphoric uses, while u-bridging patterns with contextual uniqueness.

8.4.2 The distribution of articles along the referential scale

As mentioned in Section 8.2, Dryer (1989, 2013a,b) observes that there are more definite than indefinite articles in the world’s languages. The distribution of articles in the sample of the present study could not entirely confirm this, in part because of a more fine-grained distinction of different article types. However, we find an effect in the distribution of articles along the referential scale that parallels Dryer’s findings in a way.

Distinguishing between pragmatic definite, semantic definite, specific, and nonspecific functions, we can examine the number of articles in the sample that encode these.⁶ The resulting coverage of pragmatic definite, semantic definite, specific, and nonspecific functions in the sample is shown in numbers and proportions in Table 8.7. The number of articles in the sample that expresses pragmatic definites consists of all anaphoric, definite, recognitional, inclusive-specific, and referential, i.e. $22+46+5+3+8=84$ articles. By analogy the number of articles expressing semantic definites is the sum of all definite, weak definite, inclusive-specific, and referential articles. For the specific function, all inclusive-specific, exclusive-specific, indefinite, and referential articles from the sample are counted in. The number shown for the nonspecific function corresponds to the sum of indefinite, nonspecific, and referential articles in the sample. The proportions for the four main referential functions are calculated from the total number of 141 articles in the sample.

Table 8.7: Coverage of definite, specific, and nonspecific referent types

referent types	N expressed	proportion
pragmatic definite	84	0.60
semantic definite	59	0.42
specific	62	0.44
nonspecific	51	0.36

⁶See Table 8.1 on page 296 for a summary of the distribution of article types in the sample.

The important difference between Dryer's findings and the results of this study is that the proportions in Table 8.7 do not reflect a claim about the frequency of article types in the world's languages as such. They show the crosslinguistic trend of articles encoding certain domains of referential functions. Even though definite articles are not substantially more frequent than indefinite articles in the sample (46 vs. 38), Table 8.7 reveals a trend in the distribution of articles that follows the referential scale. Most of the articles in the sample, i.e. 60%, encode pragmatic definites. The number of articles that encode semantic definite functions is lower and corresponds to 42%. Specific and nonspecific functions reach a similar number, namely 44% and 36%, respectively. These numbers show that pragmatic definites correspond to the referential functions that are most often encoded by articles, substantially more often than at least the nonspecific function. This can be formulated as a statistical universal, Universal 2, which captures a crosslinguistic trend without making predictions about single languages.

Universal 2 *The expression of referential functions by articles*

The probability of referential functions being encoded is highest for pragmatic definites and lower for semantic definite, specific, and nonspecific functions.

Thus, although definite articles are not necessarily more frequent across languages than indefinite articles, we can observe a trend in this direction when looking at the probability of referential functions being encoded by articles. In that, the results of this study are compatible with another observation of Dryer (2014), namely that anaphoric articles are more common in the world's languages than other types of articles.

8.5 A typology of articles

This section discusses the typology of articles that falls out from the data discussed in the previous chapters and from the referential scale presented in the previous section. I also compare the typology to an earlier typology of articles proposed by Dryer (2014). Despite a number of differences mostly due to different definitions of the referential concepts, both typologies converge on a number of observations.

8.5.1 Results of the present study

From the article types presented in this study and their distribution along the referentiality scale, Table 8.8 shows a typology of articles. Besides the better attested types of articles, I tentatively include weak inclusive-specific and weak referential articles in the typology. Furthermore, Table 8.8 only retains a subset of the referential functions discussed that are relevant for the distinction of article types.

Although Table 8.8 draws a synchronic picture of the types of articles attested across languages, it also offers insights to the diachrony of articles, i.e. their development and functional extension. We find a variety of articles in the definite domain. Leaving aside recognitional, weak definite and weak inclusive-specific articles for a moment, Table 8.8 shows how anaphoric, definite, inclusive-specific, and referential articles reflect various stages of functional extensions along the referential scale from pragmatic definite to semantic definite to specific and then to nonspecific.

Anaphoric articles mainly encode anaphoric referents and can be used to mark recognitional and establishing referents as well. Some anaphoric articles like the one in Akan are even able to express situational uniqueness. However, anaphoric articles are not able to encode contextually unique and u-bridging referents. Definite articles represent a further step in the development in that they cover the entire definite domain. We then find inclusive-specific articles, which can encode all definite as well as specific functions. Finally, referential articles cover the entire definite and indefinite domain; they no longer distinguish between single referential functions and instead signal that the nominal expression is a referential one.

Table 8.8: A typology of articles

article types	pragmatic definite			semantic definite		indefinite		example	
	R _{anaph}	R _{recog}	R _{establ}	R _{rel-bridge}	R _{sit.unique}	R _{cont.unique}	R _{spec}		R _{nspec}
RECOG		✓	(✓)						Oksapmin
ANA	✓	✓	✓	(✓)	(✓)				Akan
DEF	✓	✓	✓	✓	✓	✓			Kakchikel
DEF _{WEAK}					✓	✓			German
INSPEC	✓	✓	✓	✓	✓	✓	✓		Ayoreo
INSPEC _{WEAK}					✓	✓	✓		Tongan(?)
REF	✓	✓	✓	✓	✓	✓	✓	✓	Rapa Nui
REF _{WEAK}					✓	✓	✓	✓	unattested
EXSPEC							✓		Biak
INDEF							✓	✓	Bonan
NSPEC								✓	Lakota

In the indefinite domain, we find the three article types of specific, nonspecific, and indefinite articles, which express specific and nonspecific referents and the combination thereof. As was mentioned in Section 4.3, specific articles are usually thought to develop into indefinite articles, which again corresponds to an extension in the same direction on the referential scale. With the

exception of weak inclusive-specific articles, the articles that cross-cut the definite and indefinite domains encode all functions in the definite domain, i.e. pragmatic as well as semantic definites. This suggests that, indeed, the development of articles is directed from pragmatic definite > semantic definite > specific > nonspecific along the referential scale, and not attested in the opposite direction. Nonspecific articles may be viewed as an exception to this directed development, but as will be argued in Section 9.1.2, they usually originate from former indefinite articles which are restricted to nonspecific contexts in a subsequent development.

The functional domains of the weak definite and weak inclusive-specific articles can also be explained by a later restriction, blocking their use in pragmatic definite contexts (cf. Section 9.1.1). Both types of articles are very rare, and especially for the latter, more research is needed to clarify the status of the Tongan inclusive-specific article or to provide evidence for this article type from other languages. Another article type shown in Table 8.8 is the weak referential article, which would encode semantic definite, specific, and nonspecific functions. While weak referential articles are not attested to the best of my knowledge, there is no reason to categorically exclude their existence. Another potential counter-example to the functional extension of referential markers from left to right in Table 8.8 are the possessive markers discussed in Section 4.2. To a certain extent, possessive markers like the ones from e.g. Udmurt provide evidence for the development from semantic definite to pragmatic definite marker, since they can be used to mark e.g. anaphoric referents as well. However, I argued that their use in such contexts follows from other pragmatic functions these markers have and is not necessarily a primarily referential development.⁷ With regard to the directed extension of articles along the functions on the referential scale, we can thus formulate Universal 3. This universal should nevertheless be taken as a general trend and not a categorical universal that would rule out all developments in the opposite direction, since we saw that possessives can extend their functions in the opposite direction as a by-product of a different development driven by discourse-pragmatic factors.

Universal 3 *The functional extension of articles*

Articles generally extend their functions along the referential scale, from pragmatic definite to semantic definite to specific and to nonspecific contexts.

The last point of this section is an obvious one. In general, Table 8.8 highlights that articles only cover adjacent referential functions on the referential scale. Even though this statement is circular, given that the scale is derived from the functions that the attested articles have, it nevertheless makes an important prediction regarding (im)possible types of articles. Universal 4 formulates this as a restriction, categorically excluding the existence of certain article types.

⁷Moreover, as was mentioned in Section 4.2, the crosslinguistic evidence for the development of definite articles from possessive markers is not very strong; in the languages for which this development was proposed, both the definite article and the possessive can be related to a third common source like a demonstrative or a personal pronoun.

Universal 4 *Articles encode adjacent functions on the referential scale*

Articles do not encode non-adjacent functions on the referential scale. This excludes the existence of, for instance, an article that would encode definite and nonspecific but not specific functions.

8.5.2 A comparison with Dryer's typology

The most notable previous study concerned with a typology of articles is the one of Dryer (2014), who proposes a "Preliminary typology of articles". Dryer's typology is reproduced in Table 8.9 below.⁸ His choices of referential functions being somewhat different from those in the present study, I will focus on converging results as well as on the differences between the two typologies, trying to explain the latter.

As can be seen in Table 8.9, Dryer (2014) also distinguishes between largely pragmatic and semantic definite uses ("anaphoric definite" vs. "nonanaphoric definite") in the definite domain. Besides nonspecific uses (called "semantically nonspecific indefinite"), he makes an additional distinction in the indefinite domain between two uses that he calls "pragmatically specific indefinite" and "pragmatically nonspecific but semantically specific indefinite". These two types both correspond to the specific function as defined in this study, the difference being the discourse prominence of the referents. This is explained as follows by Dryer:

"For example, if someone says *I went to a movie last night*, and then goes on to talk about the movie, then a movie is normally pragmatically specific. However, if someone says the same sentence but then proceeds to say nothing more about the particular movie, then *a movie* is normally pragmatically nonspecific. Note that in both cases here, *a movie* is semantically specific, since in both cases there is an entailment that there is *a movie*." (Dryer 2014: 236-237)

This last difference is not reflected directly in the referent types in this study or in my typology shown in Table 8.8, simply because the referential functions (as Dryer also notes) are identical. The difference is nevertheless reflected in the present study in the distinction between indefinite and presentational articles (cf. Section 6.2). In contrast to Dryer (2014), I do not treat presentational articles as a subtype of specific articles because they are not necessarily restricted to be used with specific referents only. It may be the case that exclusive-specific articles, i.e. articles that strictly encode specific referents also show differences in their compatibility with less prominent referents. Such a trend was not evident in the data of the present study, but corpus studies may help to uncover more fine-grained differences regarding the use of articles in the indefinite domain and their sensitivity to discourse prominence.⁹

⁸Dryer (2014) uses abbreviated labels to refer to the 15 article types; I replaced them by numbers for the sake of simplicity.

⁹The same may be true for articles in the definite domain; especially anaphoric articles seem to differ to a greater extent in that respect.

Table 8.9: Dryer’s preliminary typology of articles

	anaphoric definite	nonanaphoric definite	pragmatically specific indefinite	pragmatically nonspecific but semantically specific indefinite	semantically nonspecific indefinite
1	Garrwa				
2		Ma’di			
3			Anufo		
4				unattested	
5					Gbeye
6	English (DEF)				
7		unattested			
8			Ngizim		
9				Siar	
10	Kokota				
11		unattested			
12			English (INDEF)		
13	Tokelauan				
14		Tz’utujil			
15	Basque				

Returning to the typology proposed by Dryer (2014) in Table 8.9, we see different article types arranged according to their functional broadness. First, we have a group of five article types (1-5) that each cover one of the five referential functions distinguished. The types 1 and 2 correspond to anaphoric (strong definite) and weak definite articles. These two types were extensively discussed in Chapter 5, and the data from the present study confirms that anaphoric articles are crosslinguistically frequent. This is not the case with weak definite articles, which will be dealt with in Section 9.1.1 (cf. also Section 5.3.5 for a discussion of the Ma’di data). The third type distinguished (exemplified by Anufo in Table 8.9) largely corresponds to what I discuss as exclusive-specific articles, with the additional distinction that the article is only used with discourse-prominent referents. As was mentioned before, such articles, called presentational articles here, do not need to be restricted to indefinite specific contexts but can be compatible with (semantically) nonspecific

ones. Such an article type, however, does not fall out from Dryer's typology; it would correspond to the two non-adjacent contexts of pragmatically specific indefinite and semantically nonspecific indefinite. As the data in the present study showed (cf. Section 6.2), we do find articles that are used in such a way, which is why the pragmatic restrictions have been kept separately from the semantic restrictions in the present study.

Being most evident in the indefinite domain, pragmatic restrictions may also apply to articles in the definite domain. This was briefly mentioned regarding anaphoric markers in Hausa and Lakota (cf. Sections 5.3.2 and 5.3.3). We saw that adnominal anaphoric markers can be restricted to certain types of topical referents. Such markers were not included as articles in the present study because they reflect the distribution of adnominal demonstratives in many languages. Not calling them articles does not mean, however, that we would not see discourse-pragmatic restrictions with referential markers in the definite domain. Going beyond the scope of the present study, such effects are probably not uncommon crosslinguistically.

The next article type (4) in Dryer's typology is unattested and corresponds to an article used with specific referents that are not discourse-prominent. Given the development of indefinite or specific articles relying on referents that are discourse-prominent, this is hardly surprising and compatible with the findings of the present study. Type 5 corresponds to nonspecific articles in the terminology used in the present study, while not very frequent crosslinguistically, the data from the present study confirms that this type of articles is attested.

Article types 6 to 9 include article types that cover two adjacent uses on Dryer's referential scale. Type 6 corresponds to definite articles, covering the entire definite domain. Type 7, used with semantically definite referent types and discourse-prominent specific referents, is marked as unattested. In the same way, Dryer (2014) proposes type 11, which also includes less discourse-prominent specific referents. Like type 7, type 11 is marked as unattested. The present study mostly parallels these findings. However, the inclusive-specific article in Tongan is a likely type 11 article in Dryer's typology.

The article types 8 and 9 largely correspond to exclusive-specific and indefinite articles in the terminology of the present study.¹⁰ The next group of articles (10-12) includes the indefinite article (type 12), the weak inclusive-specific article mentioned above (11), and type 10, exemplified by Kokota. As the source of the article in Kokota, Dryer (2014) cites Palmer (2009). Palmer (2009: 80) indeed describes the article *ia* as being used with definite and specific referents, but he also notes:

¹⁰Siar is mentioned as an example of an article that is used to encode nonspecific and specific referents, the latter not being discourse-prominent. As a source for Siar, Dryer (2014) mentions Ross (2002), the chapter on Siar in an overview volume of Oceanic languages in Crowley et al. (2002). Siar (Siar Lak) is also part of the sample of the present study based on the grammatical description of Rowe (2005). I analyse the system as having an inclusive-specific and a nonspecific article. Rowe (2005: 15-20) makes the same distinction, only that he calls the two articles referential vs. non-referential.

“Subsequent mentions of an introduced participant are typically accompanied not by an article, but by a demonstrative, indicating the definiteness of the referent, and therefore its identity as the previously mentioned participant.” (Palmer 2009: 80)

This strongly suggests that the article is not only restricted by discourse-pragmatic factors with specific referents, but that its sensitivity to such additional pragmatic factors are more global. Therefore, such restrictions are kept separate from the distinction of referential functions in the present study. I would therefore treat articles of type 10 as inclusive-specific articles as well. As was mentioned before, it is probably crosslinguistically common to restrict the use of articles to discourse-prominent referents. More work will be needed to examine the interaction between referential markers and discourse-pragmatic factors within and across languages in more detail.

The last group of articles in Dryer’s typology comprises types 13 to 15. Type 13 also corresponds to what I analyse as inclusive-specific articles. Type 15 corresponds to referential articles in the present study. An article of type 14, which is used in semantic definite, specific, and nonspecific contexts, is what I described as a weak referential article. As I mentioned above, my sample does not include any article of this type and I am not aware of any examples other than the one given by Dryer, namely Tz’utujil. Also Dryer (2014: 240) explicitly mentions that the Tz’utujil article is the only case he is aware of, explaining the situation in more detail in a footnote:

“The other two unattested types both involve combinations of nonanaphoric definites with different subsets of indefinites, but ones that do not include anaphoric definites. In fact, apart from articles that are used only with nonanaphoric definite noun phrases, my data includes only one language with an article that is used with nonanaphoric but not with anaphoric definites, namely Tzutujil.” (Dryer 2014: 240)

“The characterization of Tzutujil by Dayley (1985:255) implies that what he calls the indefinite article can be used for nonanaphoric definites as well as at least some indefinites, but it is not clear whether it can be used for all three types of indefinites, so it is not clear whether it is specifically ND + PSI + PNI + SNI [type 14, LB] rather than ND + PSI [type 7, LB] or ND + PSI + PNI [type 11, LB]. Furthermore, Tzutujil does appear to distinguish nonanaphoric definites from indefinites, in that there is a definite article that occurs alone with anaphoric definites but combines with what Dayley calls the indefinite article to mark what appears to be nonanaphoric definites.” (Dryer 2014: 240)

This confirms that articles of type 14 or weak referential articles are extremely rare at best. However, I would argue that the article in Tz’utujil is not a clear example of a weak referential article. The article in question is *jun*, presented as an indefinite article in Section 6.1.1.¹¹ According to my

¹¹These examples are from the variety of Tz’utujil spoken in San Antonio.

consultants, *jun* does not normally occur with anaphoric or contextually unique referents. In such contexts another marker *ja(r)* is used. This is shown in example (676). *Ja(r)* is usually analysed as a definite article (Dayley 1985: 254-256 and Tz’utujil Tinaamitaal 2007: 87-90).

- (676) b’anitzra k’ool wa [(ja) q’atb’al] tzijchpam jawa tinamet?
 where EXIST Q DEF town.hall inside DEM town
 ‘Where is the town hall in this town?’
 Tz’utujil (primary data)

I do not analyse the marker *ja(r)* as a definite article. As was mentioned in Section 5.1, the use of *ja(r)* never appears to be obligatory in definite contexts; according to my consultants, this may be due to individual and dialectal preferences, and the use of *ja(r)* is likely also conditioned by discourse-pragmatic factors. Its use with contextually unique referents at first mention (cf. example (676)), however, also showed that *ja(r)* can occur without *jun* in semantic definite contexts, contrary to what Dryer (2014: 240) suggests. More evidence against an analysis as a definite article comes from Dayley (1985: 255), who notes that the marker “is often used with non-third person pronouns when they are topics or Subjects [...]”. He also describes contrastive topic constructions, all of which feature *ja(r)* (Dayley 1985: 258). Thus, *ja(r)* appears to be similar to a topic marker that may be reanalysed as a definite article following the distribution of the Spanish definite article.

The status of *ja(r)* as a definite article being a separate issue, the reason why Dryer (2014: 240) analyses *jun* as an article used with both specific and semantic definite referents is probably the following example provided by Dayley (1985: 255):

- (677) [Ja jun taa7] yuuk’am to nkáaja rxin nkaxlaan.
 the a Señor carried.3 here box.my of soap.my
 ‘The (identifiable but not yet mentioned) Señor brought me a box of soap.’
 Tz’utujil (Dayley 1985: 255)

While both markers may indeed be used together in semantic definite contexts, semantic definite referents are not required to be marked in this way, as was shown in (676). If *ja(r)* is indeed a topic marker, the definite interpretation of *taa7* ‘Señor’ could also arise as an effect of its pragmatic status and not be due to the use of *jun*. This and the fact that *jun* is usually not used to mark semantic definite referents calls into question the classification of *jun* as a type 14 article.

8.6 Summary

This chapter dealt with the crosslinguistic distribution of articles. Section 8.1 was concerned with the overall distribution of the 10 major article types in the sample. Their distribution in the

crosslinguistic sample of 141 articles confirmed earlier findings that definite articles are the most frequent type of articles, accounting for 33% of the sample. However, the sample of the present study also showed that if a more fine-grained distinction of article types is applied, definite articles are not substantially more frequent than indefinite articles across languages, which make up 27% of the sample. The third most frequent article type was shown to be the anaphoric article (16%), also confirming the findings of previous studies that this article type is crosslinguistically common. The other article types were shown to be less frequent in the world's languages.

Section 8.2 then turned to the question of how the article types are distributed in different areas of the world, distinguishing between the 6 macro areas of Africa, Australia, Eurasia, North America, Papunesia, and South America. The following major trends were identified and discussed: a high proportion of definite articles in Africa and North America, a high proportion of indefinite articles in Eurasia and South America, a high proportion of exclusive-specific articles in (especially West) Africa, a lack of articles other than anaphoric and recognitional articles in Australia, and the areal restriction of recognitional articles to Australia and Papunesia. The most noteworthy macro areal finding was the absence of articles in Australia, confirming previous findings. Adnominal markers which involve identifiability were shown to have discourse-pragmatic functions, marking topicality rather than referentiality as such. This often results in adnominal anaphoric markers not being required to encode anaphoric referents.

Section 8.3 discussed the crosslinguistic absence of generic articles. By showing that various other article types are used to mark generic referents in the world's languages, I argued that the marking of generic expressions cannot be a primary function of these articles. The absence of generic articles in the world's languages was related to two factors. First, I argued that generic contexts are generally not compatible with the functions that emerging articles have. As a consequence, generic articles would only develop from articles whose use in other contexts is restricted in some way. Second, I argued that generic articles may not develop simply because the general interpretation of generic expressions is often evoked by the predicate rather than the argument (i.e. the nominal expression).

Section 8.4 revisited the referential scale on the basis of the referential of the different article types. Largely confirming the structure of earlier referential scales, this scale integrated a more fine-grained distinction of definite uses and the non-referential use. In contrast to previous studies that grouped together deictic and situationally unique contexts, the functional extension of anaphoric and weak definite articles showed that they have to be distinguished and that situationally unique contexts have to be located between pragmatic and semantic definites on the scale. In addition, I examined the distribution of the articles in the sample along the referentiality scale. I showed that we find a cline along the scale; most articles (60% of the sample) encode pragmatic definites, as opposed to nonspecific referents, marked by only 37% of articles in the sample.

Section 8.5 proposed a typology of articles, comparing it with the typology of Dryer (2014). Both typologies included various article types marking functions from either the definite or the indefinite domain. Moreover, the present study could confirm that domain-crossing article types usually encode the entire definite domain. The only exceptions to this are weak inclusive-specific and weak referential articles. I argued against the analysis of a Tz'utujil article as a weak referential article in Dryer (2014). However, given the crosslinguistic scarcity of weak definite and weak inclusive-specific articles, it is hardly surprising that weak referential articles are extremely rare or not attested at all across languages.

Chapter 9

Article systems

Chapter 9 deals with article systems and their distributions in the world's languages. In the first part (Section 9.1), I discuss two types of dependencies between articles within single languages. In both cases, I argue that this synchronic dependency can be explained by the diachronic processes that have led to the respective article systems. I then turn to cases of functional overlap between articles with single systems in Sections 9.2 and 9.3. I first describe the functional overlap between articles observed in the definite domain before examining the overlap in the indefinite domain in more detail with data from Tepehua, Basque, and Maori. The last part of this chapter, Section 9.4 proposes a typology of article systems. I show that we have to distinguish (at least) 22 types of attested article systems. Their distribution shows that simpler systems consisting of single articles are crosslinguistically more frequent than more complex systems with 2 or 3 articles. Nevertheless, the distributions also show that different types of systems prefer to divide the referential space in a certain way. The referential space can be understood as all the functions of the referential scale. A division into two domains of this space is possible and attested in two different ways: specific referents can either be grouped together with nonspecific ones, leading to a definite vs. indefinite distinction, or they can be grouped with definite referents, resulting in an inclusive-specific vs. nonspecific distinction. The crosslinguistic distributions of article systems suggest that the former is much more frequent than the latter.

9.1 Dependencies between articles

As was shown in Chapters 5 and 6, the existence of certain types of articles seems to be dependent on the existence of other types of articles within a given language. One such case is the weak definite article, which only occurs in languages that also have another anaphoric article. This dependency will be discussed in Section 9.1.1. The other type of dependency found between

articles is the one of nonspecific articles which are only attested together with exclusive-specific or inclusive-specific articles. This relation is discussed in Section 9.1.2.

9.1.1 Weak definite and anaphoric articles

Weak definite articles appear to be very rare in the world's languages and it seems that there is only one type of scenario that leads to their development. The starting point is a system with a definite article and an additional anaphoric marker. If the anaphoric marker becomes more and more conventionalized in anaphoric contexts, i.e. an anaphoric article, it may block the use of the definite article in such contexts. Then, the former definite article becomes a new weak definite article besides the anaphoric article in the system.

Such a development is also plausibly the origin for the weak inclusive-specific article in Tongan section 7.1.3. R. Clark (1974) argues that the definite accent originates from a demonstrative consisting of a vowel. The vowel was then assimilated to the final vowel of the nominal expression, leaving only a lengthened vowel which resulted in the stress shift observable in Modern Tongan. Thus, also in the case of Tongan, it is likely that a former demonstrative developed into a systematic anaphoric article, eventually taken over the use of the inclusive-specific article in (most) pragmatic definite contexts.

The only alternative scenario would involve the development of a (weak) definite article from a possessive marker. Such a development is very implausible for two reasons. First, as was discussed in more detail in Section 4.2, there is no strong crosslinguistic evidence for the development from a possessive marker (in the strict sense) to a definite article. Second, possessive markers that develop into some kind of definite marker, as in e.g. Udmurt, are also found in anaphoric contexts. In other words, they are not restricted to semantic definite contexts but are used in pragmatic definite contexts as well.

From the data presented in Section 5.3, another tendency regarding weak definite articles emerged, namely that the availability of an anaphoric article or marker does not necessarily block the use of the functionally broader definite article in pragmatic definite contexts. I showed for Hausa, Lakota, and Ma'di that the presence of an additional anaphoric marker does not restrict the definite article to semantic definite contexts.¹ The data from Urama (cf. Section 5.3.4), another language from the sample with an anaphoric and a definite article, provided more evidence for the co-existence of a broader definite and the semantically more restricted anaphoric article.

Another language from the sample similar to Urama is Jamsay. Jamsay is a Dogon language, spoken in the regions around the border between Mali and Burkina Faso by over 30,000 speakers (Heath 2008: 2). Jamsay has a definite article, *kuⁿ*, but it can also use the anaphoric marker *kò* in

¹The alleged weak definite article in Hausa is in fact an anaphoric (and not definite) article (cf. Section 5.3.2).

addition to the definite article with anaphoric referents. Again, the availability of an anaphoric marker in Jamsay does not necessarily restrict the definite article to semantic definite contexts. In fact, the anaphoric marker is often used together with the definite article in this language (Heath 2008: 164). Examples (678) and (679) show the use of the definite article *kùⁿ* with a contextually and a situationally unique referent.

(678) [àná kùⁿ fú:] bérè
 village ART:DEF all in
 ‘in all the villages’
 Jamsay (Heath 2008: 282)

(679) ù-jùwⁿó kò [ñǎ: kùⁿ] jé méy [gǔ:n kùⁿ] dìgé nú:-yⁿè
 RED-mouse DEM food ART:DEF for and bowl.cover ART:DEF follow enter-PFV.S:3SG
 tánjà: dèy, ...
 happen if ...
 ‘The mouse, when it has followed along the bowl cover and has gone in (under the bowl itself) to get the food, ...’
 Jamsay (Heath 2008: 302)

In anaphoric contexts, example (680) shows how the definite article can be reinforced by the anaphoric marker *kò*.

(680) tògù pó:ró tógó kó bèrè: bè nú: [kò tóg kùⁿ] úró
 shed first shed.building 3SG.NHUM in 3PL enter.PFV ART:ANA shed ART:DEF house
 táná-ńá mèyⁿ nîŋ yó=kò.
 become-CAUS and now exist=NHUM
 ‘The first shed that they built and went into (to live), that (aforementioned) shed having been transformed into a house, it is still there to this day.’
 Jamsay (Heath 2008: 164)

Now, one could also analyse *kùⁿ* as a weak definite article and the combination of *kò ... kùⁿ* as the anaphoric article. However, the use of *kò* is not required in anaphoric contexts. This is shown in (681a). Moreover, the definite article *kùⁿ* occurs in other pragmatic definite contexts, e.g. to establish referents as in (682). Therefore, it does not behave like the weak definite articles in German or Fering whose use in such contexts is infelicitous.

(681) a. úrà yěy-yà-bà déy nè, tàrá íjé nè iñé=yⁿ ...
 house.LOC come-PERF-S:3PL if now coll.hunt today now what=FOC ...
 ‘When they [hunters] have come how, (from) the collective hunt now, what do they do with them [animals]?’

- b. ùrò-dú: mà bèrê:, ìnè gàmà-nám yó=wò-bà, bé sày, bé
 family POSS inside person certain-PL exist=be.human-s:3PL 3PL only POSS:3PL
 ñà:-bà:ñá sày, [tàrá kùⁿ] ìnè dó:-n sà:-rá-bà.
 meal-bowl only coll.hunt ART:DEF person arrive.IPFV-POSS:PL.SG have-NEG-s:3PL
 ‘Within the family, there are certain people, only they and their (wooden) food bowls;
 they (perhaps) do not have anyone who (can) go on the collective hunt.’
 Jamsay (Heath 2008: 701-702)

- (682) ijù gô: kùⁿ bé
 dog go.out.PERF ART:DEF PL
 ‘the dogs who went out’
 (Heath 2008: 255)

While a clearer picture would require further work on the distribution of the two markers in the individual languages, the trend emerges that the anaphoric article or marker in such systems is often additionally restricted pragmatically. It is typically used to signal a topic shift, or to re-activate discourse referents whose antecedent is further away.

Also the definite article in Kaqchikel (cf. Section 5.1.1) was shown to be blocked by a stronger anaphoric demonstrative in certain cases. Examples (683) and (684) show contexts in which the anaphoric demonstrative *ri ... ri* is required instead of the definite article *ri*.

- (683) a Lu xiroyoj jun b’eychik mambajota yich’o ruk’in [ri vinaq
 PROP Pedro call.PST.s:3SG.o:1SG one way want.NEG.1SG talk with.3SG ANA person
 *(ri)].
 ANA
 ‘Pedro called again. I don’t want to talk to that person.’
 Kaqchikel (primary data)

- (684) k’ab’a a Lu xuroyoj ri ru vecino loman chuk’a
 recently PROP Pedro call.PST.s:3SG.o:3SG ART:DEF POSS:3SG neighbour while also
 ink’o yin chila’ [ri achin *(ri)] jabey xqa chinwach.
 be.1SG I there ANA man ANA nice came in.front
 ‘Recently Pedro invited his neighbour while I was at his house. I liked that man.’
 Kaqchikel (primary data)

Although it was shown in section 5.1.1 that the definite article in Kaqchikel occurs in anaphoric contexts as well, examples like (683) and (684) show that its use can be blocked in certain types of anaphoric contexts. Note that the use of a stronger anaphoric marker is paralleled in English; the demonstrative *that* seems much more appropriate than the definite article in the contexts of (683) and (684).

Thus, non-anaphoric definite articles are extremely rare across the languages of the world. To the best of my knowledge, the attested cases in which an anaphoric article properly blocks the

use of the definite article in pragmatic definite contexts are mostly confined to different Germanic varieties and may be restricted to certain types of constructions. Icelandic, for instance, has a definite article (suffixed to the noun) that is used with pragmatic as well as semantic definites by default. Only in the presence of an evaluative adjective in the nominal complex, a separate weak definite article, a pronominal free form, is used with semantic definites, while being infelicitous in pragmatic definite contexts (Ingason 2016: 115-125). Example (685) shows this with a unique definite in (685a) and an anaphoric definite in (685b). Without the presence of an evaluative adjective, however, the regular definite article is used in both types of contexts, as can be seen in (686).

- (685) a. Tim Berners Lee kynnti heim-inn fyrir [hinum ótrúlega
Tim Berners Lee introduced world-ART:DEF to ART:DEF_{WEAK} amazing
veraldarvef].
world.wide.web
[Context: first mention of the world wide web.]
'Tim Berners Lee introduced the world to the amazing world wide web.'
- b. Hún fékk engin góð svör frá [*hinum hræðilega stjórnámálamanni].
she got no good answers from ART:DEF_{WEAK} terrible politician
[Context: Mary talked to a writer and a terrible politician.]
'She got no good answers from] the terrible politician.'
Icelandic (Ingason 2016: 123, 133)
- (686) a. ... veraldarvef-**num**.
... world.wide.web-ART:DEF
[Context: first mention of the world wide web]
'[Tim Berners Lee introduced the world to] the world wide web.'
- b. ... stjórnámálamanni-**num**.
... politician-ART:DEF
[Context: Mary talked to a writer and a politician.]
'[She got no interesting answers from] the politician.'
Icelandic (Ingason 2016: 118)

The Icelandic examples, together with the German and Fering data presented in Section 5.3 show that article systems do exist where a definite article is restricted to semantic definite contexts by the presence of another anaphoric article, at least in certain contexts. However, the data presented from Hausa, Lakota, Urama, Ma'di, and Jamsay suggest that this is usually not the case. Thus, the availability of an anaphoric article often does not make the use of the definite article in anaphoric and other pragmatic definite contexts infelicitous.

9.1.2 Nonspecific and specific articles

The sample includes 5 languages with nonspecific articles, namely Biak (Section 6.3.1), Tongan (Section 6.4.3), Lakota (Section 6.4.2), Ayoreo (Section 6.4.1), and Siar Lak. Biak and Lakota additionally have a definite and an exclusive-specific article. Siar Lak, Tongan, and Ayoreo all have an inclusive-specific article that is opposed to the nonspecific article. To illustrate this distribution across the referential space, examples (687) and (688) show the use of the Siar Lak inclusive-specific article *ep*. Opposed to that, the nonspecific article *ti* is used in (689) to mark the referent as nonspecific.

(687) a. Ap i pastat pas i tik ep sói, ep móróu.
 and 3SG find COMPL 3SG one ART:INSPEC snake ART:INSPEC kind.of.snake
 ‘And he found a snake, a *móróu*.’

b. Ap e Tagorman i tasim akak o-n [ep sói] ...
 and PROP Tagorman 3SG know good OBL-3SG ART:INSPEC snake ...
 ‘And Tagorman knew **the snake** well ...’
 Siar-Lak (Rowe 2005: 105)

(688) I ding ep kirai diat sin pastat pas i tik [ep pun] i
 3SG that ART:INSPEC day 3TRI sibling find COMPL 3SG one ART:INSPEC turtle 3SG
 sót.
 come.ashore
 ‘That day the three brothers found a **turtle** that came ashore.’
 Siar-Lak (Rowe 2005: 50)

(689) Na ep wang el buh [ti yai] o [ti at] ...
 if ART:INSPEC canoe 3SG.POT hit ART:INSPEC tree or ART:INSPEC stone ...
 ‘If the canoe hits a **tree** or a **stone** ...’
 Siar-Lak (Rowe 2005: 18)

As was shown in Section 6.4.4, many of the nonspecific articles (or markers) in North and Central American languages probably go back to a former indefinite article, which is combined with or replaced by an irrealis marker in nonspecific contexts. Thus, (690) shows for Hidatsa that the indefinite article *wa* is only used in specific contexts. If the referent is nonspecific, (690b) shows that it is expressed as an irrealis predication using the irrealis marker *rúg* with the nominal expression. The same can be seen in (691) for Q’anjob’al, only that the irrealis marker *-oq* attaches to the (former) indefinite article *jun* instead of replacing the article in nonspecific contexts.

(690) a. hiraacá-mià-wa m-úá-waa-c.
 Hidatsa-woman-ART:INDEF 1-marry-CAUS.1-DECL
 ‘I married a Hidatsa woman.’

- b. hiraacá-mià-rúg aru-m-úá-waa-c.
 Hidatsa-woman-COND IRR-1-marry-CAUS.1-DECL
 ‘I am going to marry a Hidatsa woman.’
 literally: ‘If she is a Hidatsa woman I will marry her.’
 Hidatsa (Park 2012: 368)

- (691) a. man anima-oq hach.
 NEG person-IRR 2SG
 ‘You are not a person.’
 Q’anjob’al (Mateo Toledo 2017: 551)
- b. ta’ watz’ max hot mulnaj hab’a ay mi wal [junoq ha-mimam na].
 if good PST COND work.2SG 2SG EXIST for.2SG EMPH ART:NSPEC 2SG-big house
 ‘If you had worked well, you would have had a big house.’
 Q’anjob’al (primary data)

The nonspecific articles from the Oceanic languages discussed in this study have been argued to go back to an indefinite article **ta* in (Western) Proto-Oceanic (Crowley et al. 2002: 71, Ross 1988: 357-360). Kieviet (2017a: 242) also mentions the reconstructed indefinite article **sa* for Proto-Polynesian. Thus, also for the Oceanic languages the most plausible source of the nonspecific article was a previous indefinite article whose use became restricted to nonspecific contexts. In contrast to the scenarios described for the American nonspecific markers, no new additional nonspecific marker seems to have developed in Oceanic. Instead, the restriction of a former indefinite article may have been the result of a functional extension of a former definite article to specific contexts. At least for the inclusive-specific articles in Tongan (*he*) and Maori (*te*), it has been shown that they go back to the reconstructed definite article **te* in Proto-Polynesian (Kieviet 2017a: 237, 242).² Thus, the definite articles in these languages could have extended their use to specific referents, replacing the use of the former indefinite article and restricting it to nonspecific contexts. How these inclusive-specific articles developed, if and how they could take over the coding of specific referents are issues that go beyond the scope of the present study.

To conclude, the important point regarding the dependency of nonspecific articles on exclusive/inclusive-specific articles is that nonspecific articles generally originate from indefinite articles whose use is restricted to or altered in nonspecific contexts. In case the article is modified by an additional marker in nonspecific contexts, its use with specific referents is continued in specific contexts, and we see the article as a new exclusive-specific article (as in Q’anjob’al). In systems with inclusive-specific articles, on the other hand, the use of the former indefinite article is rather restricted to only marking nonspecific referents (as in Tongan). In the absence of evidence of other lexical or grammatical sources for nonspecific articles, the scenarios presented in this section not

²This also holds for the referential article *te* in Rapa Nui.

only account for their dependency on specific articles in the system, but it may also explain their rare occurrence across languages, given that their development requires different rare processes to take place.

9.2 Functional overlap of articles in the definite domain

What has not been addressed so far are systems with articles that overlap in their referential functions. In the definite domain, we find this type of overlap between definite and anaphoric as well as definite and recognitional articles.

The sample contains three languages with a definite article and a functionally more restricted anaphoric or recognitional article. This is a rather conservative count, given that I do not include systems with a definite article and an additional anaphoric markers like Lakota or Jamsay, which are probably very common crosslinguistically. The three languages with two overlapping articles in the definite domain are all languages from Papunesia: Oksapmin and Lavukaleve with a definite article and a recognitional article (cf. Sections 5.4.1 and 5.4.2), and Urama with a definite and an anaphoric article (cf. Section 5.3.4). This crosslinguistic distribution points again towards an areal bias regarding the availability of specialized markers in the definite domain.

As was shown in detail in Section 5.3.4 for Urama, the anaphoric article does not necessarily fully replace the functionally broader definite article in pragmatic definite contexts. While it may seem less obvious how those articles can be used systematically and co-exist within systems, the situation is relatively similar to the competition that we find in other languages between the definite article and the demonstrative system. For instance, also in English we find contexts in which a definite article could be replaced by a demonstrative without any major semantic or pragmatic changes, as in example (692). The context of such an utterance could be a discourse situation between two interlocutors in a room with a single shelf, visible and thus identifiable to both of them.

- (692) a. Just give [**the** shelf] a quick wipe, will you, before I put this vase on it.
b. Just give [**that** shelf] a quick wipe, will you, before I put this vase on it.
(Lyons 1999: 3)

In a system like Urama, where the definite article usually co-occurs with the anaphoric article in anaphoric contexts, we saw that the definite article could be used alone as well to mark anaphoric referents as identifiable, and discourse-pragmatic factors as well as the distance to the antecedent then determine which marker is used. The alternation between the use of the definite article or the combination of definite and anaphoric article in Urama is repeated here in example (693) and (694). The combination of both articles, *aro'o ... =i* is used in both examples, while the definite

article =i alone is used in (694) to refer back to *goto'a* ‘coconut tree’ mentioned in the previous sentence.

- (693) Nu'a huna ata Iroroma vati kekai ta; [aro'o nu'a=i] modobo ka pe
 tree big some Iroroma place near LOC ART:ANA tree=ART:DEF can PRS canoe
 ededeai ri.
 make COMP
 ‘There’s a big tree near Iroroma’s place; **that tree** could make a canoe.’
 Urama (J. Brown et al. 2016: 22)

- (694) a. Ioro ohu=i tabo kiaukia bomo gema=i ro go'ota=i ahiai ka.
 climb top=ART:DEF LOC enough pig big=ART:DEF NOM coconut=ART:DEF cut PRS
 ‘He climbed to the top, and then the big pig started cutting the coconut tree down.’
- b. Go'ota=i p-ahiai ita [aro'o go'ota=i] omo'ai ri
 coconut=ART:DEF REMPST-cut then ART:ANA coconut=ART:DEF fall COMP
 a'ai ta, Iroroma imumuai ka go'ota ata=i to.
 NEARFUT LOC Iroroma fly PRS coconut other=ART:DEF ALL
 ‘He cut that coconut tree down and when it was about to fall, Iroroma flew onto another coconut tree.’
 Urama (J. Brown et al. 2016: 89)

The situation in Oksapmin and Lavukaleve with a definite article and a functionally much more restricted recognitional article is somewhat different. Based on the data provided in Loughnane (2009) for Oksapmin and Terrill (2003) for Lavukaleve, it is unclear whether or not the definite article can also occur in recognitional contexts. In addition, we saw in Section 5.4.2 that the use of the recognitional article in Lavukaleve seems especially preferred for culture-specific referents. Therefore, it is not entirely clear whether or not the two articles functionally overlap in a strict sense or whether the overlap needs to be understood in a more abstract way, meaning that the article system includes a functionally broader and a functionally more restricted article in the definite domain. Again, the situation in such systems is similar to the competition between definite articles and demonstratives in other languages. In English, for instance, both can be used equivalently in recognitional contexts, as is shown in (695b):

- (695) a. Do you remember [**the** dog] (we used to have)?
 b. Do you remember [**that** dog] (we used to have)?

Thus, even though functionally overlapping articles in the definite domain appear to be crosslinguistically rare, they fit well into the bigger picture of the competition between (different types of) demonstratives, the definite article, but also pronominal or zero forms.

9.3 Functional overlap of articles in the indefinite domain

The other type of systems with functionally overlapping articles attested in the sample all involve a functionally very broad article that cross-cuts the definite and indefinite domains, and another article that is used to disambiguate between these two domains in contexts in which such a distinction is important. There are three languages in the sample with such systems. Tepehua and Basque both have article systems with a referential article and an additional presentational article. Maori, on the other hand, has a complex system of an inclusive-specific article besides a presentational article and a predicate marker used to mark referents in the indefinite domain. This section will describe those three systems in detail. The main pattern that emerges is that the additional presentational articles are used two types of contexts. On the one hand, they occur to signal that the referent is not identifiable by the hearer when the referent is likely to be otherwise interpreted as definite based on its discourse prominence. On the other hand, the presentational articles are used to explicitly signal that the speaker is not able to fully identify the referent either. In both types of situations, inclusive-specific or referential articles would in principle be able to express such referential functions as well, but their ambiguity between a definite or specific interpretation of the referent makes them less felicitous to highlight the non-identifiability of the referent.

9.3.1 Tepehua

Huehuetla Tepehua is a Totonacan language spoken in Mexico in the town of Huehuetla and its surrounding areas. Kung (2007: viii) estimates that the language has fewer than 1500 speakers. In addition to its referential article *juu*, Huehuetla Tepehua (henceforth: Tepehua) uses the presentational article *tam*.

Although *juu* is analysed as a definite article in Kung (2007: 385), its use in definite, specific, and nonspecific contexts motivates its analysis as a referential article. Example (696) shows the use of *juu* with the anaphoric referent of *barda* ‘wall’. Note that when *barda* is introduced first in (696a) as a specific referent, it is also marked by the article *juu*.

- (696) a. 7entons nii paastak-lich juu 7ukxtin nii ka-nawii-ya7 juu barda.
then COMP think-PFV ART:REF boss COMP IRR-make-FUT ART:REF wall
‘Then the mayor thought that he was going to build a wall.’
- b. puus 7ulaa-kaa-lich [juu qex].
well put-S:INDEF-PFV ART:REF wall
‘Well, they built the wall.’
- Tepehua (Kung 2007: 673)

In (697), *juu* is used with the contextually unique referent of *laqachaqan* ‘town’. The referent is unique because the town is unique in the larger context of the discourse situation.

(697) 7an-li-ch x-st'aa-nVn-ta [juu laqachaqan].
 go-PFV-already PST-sell-O:INDEF-PFV ART:REF town
 'He went selling in the town.'
 Tepehua (Kung 2007: 643)

We also find the referential article with establishing referents that are made unambiguously identifiable to the hearer only in the immediately following discourse segment. One such example is given in (698) below:³

(698) maa xta7amaqpanan [juu papa7-nin juu kaa waa lakak'iwin
 EVID wash.clothes.PST.S:3PL.O:INDEF ART:REF man-PL REL EPIST FOC woods
 xtat'ajun].
 live.PST.S:3PL
 'The men that were living in the woods would wash.'
 Tepehua (Kung 2007: 590)

Thus, the previous examples showed that *juu* in Tepehua is used systematically in pragmatic and semantic definite contexts. As a referential article, *juu* is also used systematically with specific and nonspecific referents. Example (699) shows a prototypical context of the introduction of new specific discourse referents. We see that *juu* is used in this context as well.

(699) 7alin-li laqa-tam 7a-wilchan maa soq ta-laa-lhiitajuu [juu
 there.is-PFV CL-ART:PRES CL:ANOTHER-day EVID straight S:3PL-RECIP-find.PFV ART:REF
 7akumwarii] juu laka-x-chaqa7-7an.
 friend ART:REF PREP-POSS:3-house-POSS:PL
 'One day, two friends met in their houses.'
 Tepehua (Kung 2007: 640)

In (700) and (701), the referents of *lhiich'alhkat* 'job' and *serrootii* 'saw' are particular referents evoked by the nominal expressions without being unambiguously identifiable by the hearer. In these two cases, the identity of the referent is simply not relevant in the given discourse situations. Both nouns are marked by the article *juu*.

(700) t'asa-ni-kan-lich nii ka-xtaq-ni-kan-a7ch [juu lhiich'alhkat].
 yell-DAT-S:INDEF-PFV COMP IRR-give-DAT-S:INDEF-FUT ART:REF job
 'They yelled that they were going to give him a job.'
 Tepehua (Kung 2007: 463)

³Discussing relative clauses in Tepehua, Kung (2007: 589) mentions that "[t]he relativizer *juu* is homophonous with the definite [here: inclusive-specific, LB] article *juu*". It is plausible to assume that the relativizer *juu* is not only homophonous but diachronically related to the article *juu*.

- (701) waa ki-jun-ni-li juu lijuuntuu mim-pay nii naa qoxich [juu
 FOC O:1-say-DAT-PFV ART:REF deceased POSS:2-father COMP EMPH good ART:REF
 serrootii].
 saw
 ‘Your deceased father told me that it was a good saw.’
 Tepehua (Kung 2007: 601)

The following examples (702) and (703) show the use of *juu* with nonspecific referents. The use of *juu* with definite, specific, and nonspecific referents thus motivates its analysis as a referential article.

- (702) pero nii xa-nii-y-ch mati7-ch [juu x-lhiich’alhkati7an]
 but COMP PST-die-IPFV-ALREADY nothing-ALREADY ART:REF POSS:3-work-POSS.PL
 xa-jun.
 PST-be.IPFV
 ‘But after she was dead, there was no more work.’
 Tepehua (Kung 2007: 693)

- (703) waa jaantu-ch laa-y x-ta-lhiitajuu-y juu 7anuu [juu
 FOC NEG-ALREADY CAN-IPFV PST-3PL.SUB-find-IPFV ART:REF HESIT ART:REF
 lhiich’alhkati].
 work
 ‘They played (instruments) because they couldn’t find work.’
 Tepehua (Kung 2007: 690)

As expected, if a nominal expression is used as a predicate, *juu* is not used. This is shown in (704). In equative constructions as in (705), on the other hand, where the nominal expression *maqtili7* ‘devil’ is referential, the article occurs with the noun.

- (704) pero juu 7anu7 lapanak juu x-xaqa=lhii7an jaantu qoxiyaa lapanak waa.
 but ART:REF DEM person REL PST-pull=take.IPFV NEG good person FOC
 ‘But that person who was pulling them along is not a good person.’
 Tepehua (Kung 2007: 562)

- (705) yuuch [juu maqtili7]?
 3SG ART:REF wild.animal
 ‘Is he the devil?’
 Tepehua (Kung 2007: 565)

In addition to the referential article *juu*, Tepehua has the presentational article *tam*. Kung (2007: 385-387) introduces *tam* as an indefinite article which goes back to the numeral ‘one’. Usually, the numeral *tam* is used together with a nominal classifier. Kung (2007: 386-387) notes that if used

as a marker of indefiniteness, the presence of a classifier is not required (although possible). She further notes that if *tam* is used as the numeral ‘one’, it occurs together with the referential article *juu* as in (706).

- (706) maa x-kitasp’it’-ta-ch x-7a-st’aa-nVn-ta [juu puma-tam
 REPORT PST-return-PFV-ALREADY PST-PL-sell-O:INDEF-PFV ART:REF CL:HUMAN-one
 x-kumwarii] laqa-tam laqachaqaan.
 POSS:3-compadre CL:GENERAL-ART:PRES town
 ‘One friend had returned from selling in a town.’
 Tepehua (Kung 2007: 387)

When used as a referential marker, *tam* can occur with both specific and nonspecific referents if the referents and the fact that they are not (yet) identifiable are of high relevance in the discourse. Therefore, I treat *tam* as a presentational article.

Examples (707) and (708) show a specific and a nonspecific context each; in both cases, the referent in question is marked by *tam*. In (707), the referent of *kin-tata7* ‘old man’ in is introduced as the protagonist of the narrative.

- (707) Maa chunch nawii-ta [puma-tam kin-tata7] ...
 EVID thus do-PFV CL:HUMAN-ART:PRES POSS:1-old.man ...
 ‘That’s what an old man did ...’
 Tepehua (Kung 2007: 631)

- (708) [tam maqaali7] ka-maa-ch’ixtaq-ninch juu tuumiin aantu qoxiyaa tuumiin
 ART:PRES rich.person IRR-CAUS-loan-DAT-O:2 ART:REF money NEG good money
 palata.
 better
 ‘A rich person could loan you money, but it isn’t good money.’
 Tepehua (Kung 2007: 616)

The context in (709) is another example of a referent that is not identifiable (and whose exact properties are irrelevant) but important in a narrative. Example (709a) introduces a nonspecific referent *maaskaraa* ‘mask’ marked by *tam*. Example (709b) shows a later utterance of the same narrative, informing the hearer that the protagonist indeed bought a mask. The referent of *laqpuu-tanuti* ‘mask’ is marked again by the presentational article *tam*. Even though the exact properties of the mask are not relevant in the narrative (or rather, they are never mentioned), the mask itself has an important role in the denouement at the end of the story, helping the protagonist to become rich.⁴

⁴The lexeme *maaskara* is most likely a loan from Spanish *máscara* ‘mask’, whereas *laqpootanuuti* is a Tepehua expression, segmentable as *laqpuu-tanuu-ti* ‘face-put.on-NMLZ’ (Kung 2007: 379).

- (709) a. ni maa tasp'it-choqo-cha min-li-ch maa tamaju-nV7
 COMP REPORT return-REP-DIST.PFV come-PFV-ALREADY REPORT buy-INF
 [laqa-tam x-maaskaraa] ...
 CL:GENERAL-ART:PRES POSS:3-mask ...
 'And when he returned again, he came to buy a mask ...'
- b. lhii-tamaw-li [laqa-tam x-laquutanuti] nii maa
 APPL-buy-PFV CL:GENERAL-ART:PRES POSS:3-mask COMP REPORT
 min-choqo-li-ch juu laka-tii.
 come-REP-PFV-ALREADY ART:REF PREP-road
 'And he bought himself a mask and he returned along the road.'
 Tepehua (Kung 2007: 647-648)

The context of example (710) is the introduction to a narrative about the history of Huehuetla. The story begins by setting the scene and listing various places where people would come from in the past to live in Huehuetla. Then, (710) introduces the event that led to the development of the story told. Even though the exact point in time of that event, an illness, is probably not identifiable (and not even relevant), the event itself is highly relevant for the remainder of the narrative. Therefore, we see that both referents of *wilhchan* 'day' and *taqanqati* 'sickness' are introduced by the presentational article *tam*.

- (710) a. pero maa min-li [laqa-tam wilhchan] juu maa naa
 but REPORT come-PFV CL:GENERAL-ART:PRES day ART:REF REPORT EMPH
 lhuu nii-li juu lapanak.
 many die-PFV DEF:REF people
 'But there came a day [when] many people died.'
- b. maa waa min-li [laqa-tam maa taqanqati] maa waa
 REPORT FOC come-PFV CL:GENERAL-ART:PRES REPORT sickness REPORT FOC
 laka-p'uch'i-li juu lapanak.
 body-rot-PFV ART:REF people
 'There was an illness, the bodies of the people rotted.'
 Tepehua (Kung 2007: 659)

Although not mentioned explicitly, the examples in Kung (2007) are suggestive of *tam* being used exclusively with singular nouns, while the referential article *juu* is compatible with both singular and plural expressions without formally expressing number (cf. example (698)). Thus, *tam* appears to be used to emphasize that there is a single, salient referent whose exact identity is either irrelevant or unknown (also to the speaker) but which is highly relevant for the following discourse.

9.3.2 Basque

Basque, an isolate language spoken in Spain and France, has approximately 700,000 speakers according to Hualde & Ortiz de Urbina (2003: 3). Traditionally, Basque is described as having a definite and an indefinite article (cf. Trask (2003: 119), de Rijk (2008: 17), and Manterola (2007)). However, already Trask (2003: 119) mentions that “[t]he label ‘definite article’ is misleading, since this article is of much broader use than the English definite article.” He notes that the so-called definite article *-a* can occur with definite, specific, nonspecific, and generic expressions amongst other uses. In a similar vein, Dryer (2014) classifies Basque *-a* as covering all referential functions distinguished in his typology (cf. Section 8.5.2). This makes the article a referential article in the terminology used in the present study. The following paragraphs will briefly give an overview of the use of *-a*, before turning to the presentational article *bat*.

Examples (711) to (713) show the use of the referential article *-a* in different types of definite contexts.⁵ In (711) and (712), we see the referential article used to mark a referent as situationally and contextually unique, respectively. In (713), *-a* marks an establishing referent, which is identifiable based on the information provided within the same utterance.

(711) Nire lagun-a gure herri-ko alkate-a da.
my friend-ART:REF our town-REL mayor-ART:REF is
‘My friend is the mayor of our town.’
Basque (Hualde & Ortiz de Urbina 2003: 180)

(712) Presidente-a-k dimititu du.
president-ART:REF-ERG resign AUX
‘The president has resigned.’
Basque (Hualde & Ortiz de Urbina 2003: 181)

(713) Hilabetekari horrentzat moldatu ditudan [lan purruxk-a-k]!
monthly that.for prepared AUX.COMP work little-ART:REF-PL
‘**The little works** that I’ve done for that monthly publication!’
Basque (Hualde & Ortiz de Urbina 2003: 568)

Examples (714) and (715) show *-a* marking a specific referent; in this case, the speaker has a particular single referent in mind, but the referent is not identifiable by the hearer.

(714) [Kotxe berri-a] erosi dut.
car new-ART:REF buy AUX
‘I’ve bought a **new car**.’
Basque (Hualde & Ortiz de Urbina 2003: 119)

⁵All of the Basque examples in this section are taken from the grammar of Hualde & Ortiz de Urbina (2003). This grammar is a collaborative effort of nine different contributors. For the sake of simplicity, I only refer to the grammar as Hualde & Ortiz de Urbina (2003) when indicating the sources of the examples.

(715) Emazte-a dut.
 wife-ART:REF have
 ‘I have a wife.’
 Basque (Hualde & Ortiz de Urbina 2003: 119)

The referential article is also used to mark specific referents in typical existential constructions. Examples (716) and (717) show this use of *-a*.

(716) Lekuederr-a-k dauda Bizkaian.
 place.beautiful-ART:REF-PL are Bizkaia.LOC
 ‘There are beautiful places in Bizkaia.’
 Basque (Hualde & Ortiz de Urbina 2003: 120)

(717) Ogi-a dago mahaian.
 bread-ART:REF is table.LOC
 ‘There is bread on the table.’
 Basque (Hualde & Ortiz de Urbina 2003: 120)

The following two examples, (718) and (719), are used as examples of nonspecific contexts in Hualde & Ortiz de Urbina (2003: 120). One could argue that the referents are specific in that they are within the scope of an episodic event and their existence is presupposed. However, since *artoa* ‘(some) corn’ and *sagarrak* ‘(some) apples’ refer to a set of referents where the exact identity of the single referents is irrelevant, they may indeed be constructed as nonspecific in that the speaker does not have any particular referent in mind. Indeed, these two contexts are similar to certain contexts that nonspecific articles were shown to occur in as well (cf. Section 6.4).

(718) Arto-a erein dugu.
 corn-ART:REF plant AUX
 ‘We have planted corn.’
 Basque (Hualde & Ortiz de Urbina 2003: 120)

(719) Sagarra-a-k jaten ari dira.
 apple-ART:REF-PL eat.NOM.LOC AUX are
 ‘They are eating apples.’
 Basque (Hualde & Ortiz de Urbina 2003: 120)

In contexts of negation or questions scoping over the referent in question, Basque often uses the partitive marker *-rik* similarly to the NPI *any* in English. It is also classified as a polarity item in Hualde & Ortiz de Urbina (2003: 124). The use of *-rik* is shown in (720), (721), and (722).

(720) Gaur ez da hemen ume-rik jaio.
 today NEG AUX here baby-PART be.born
 ‘No baby has been born here today.’
 Basque (Hualde & Ortiz de Urbina 2003: 124)

(721) Behar dugu ogi-rik?
 need AUX bread-PART
 ‘Do we need any bread?’
 Basque (Hualde & Ortiz de Urbina 2003: 125)

(722) Nola eros auto-rik diru-rik gabe?
 how buy.RAD car-PART money-PART without
 ‘How can one buy a car without any money?’
 Basque (Hualde & Ortiz de Urbina 2003: 125)

Interestingly, Hualde & Ortiz de Urbina (2003: 126) mention the minimal pair between *-rik* and the article *-a* shown in (723). If the partitive marker is used as in (723a), the referent is interpreted as nonspecific, whereas the use of the referential article as in (723b) marks it as definite.

(723) a. erantzun-ik eman baino.lehenago ...
 answer-PART give before ...
 ‘before giving any answer ...’
 b. erantzun-a eman baino.lehenago ...
 answer-ART:REF give before ...
 ‘before giving the answer ...’
 Basque (Hualde & Ortiz de Urbina 2003: 126)

The classification of *-a* as a referential article is based on the fact that *-a* does nevertheless occur with nonspecific referents. In example (724), *zigarro* ‘cigarette’ is nonspecific with no particular referent being evoked according to Hualde & Ortiz de Urbina (2003: 120). Although not mentioned explicitly as containing nonspecific referents, the contexts in (725) and (726) involve a nonspecific referent by default as well. In both cases, we see that the referential article *-a* is used.

(724) Zigarro-a nahi dut.
 cigarette-ART:REF want AUX
 ‘I want a cigarette.’
 Basque (Hualde & Ortiz de Urbina 2003: 120)

(725) gizon-a-k irabaz eta bere baleza ere mundu guzi-a.
 man-ART:REF-ERG win and his if.AUX also world whole-ART:REF
 ‘even if a man won and took hold of the entire world’
 Basque (Hualde & Ortiz de Urbina 2003: 272)

(726) Diru-a edukiz gero, baserri-a erosiko nuke.
 money-ART:REF have.INSTR after country.house-ART:REF buy.FUT AUX
 ‘If I had money, I’d buy a country house.’
 Basque (Hualde & Ortiz de Urbina 2003: 744)

Having established that *-a* occurs with definite, specific, and nonspecific referents, example (727) shows that *-a* cannot be used with a nominal expression if the latter is non-referential, e.g. expressing a predicate. In (728), on the other hand, we see an identificational or equative construction in which the nominal expression *iraskaslea* ‘a/the teacher’ is referential. As expected, the referential article is used in this case.

(727) Miren irakasle-(*a) dago.
 Miren teacher-ART:REF is
 ‘Miren is working as a teacher.’
 Basque (Hualde & Ortiz de Urbina 2003: 434)

(728) Jon irakasle-a da.
 Jon teacher-ART:REF is
 ‘Jon is a/the teacher.’
 Basque (Hualde & Ortiz de Urbina 2003: 432)

Thus, like the referential article in Tepehua, Basque *-a* systematically marks definite, specific, and nonspecific referents and is absent in contexts in which the nominal expression is a predicate. As was mentioned at the beginning of this section, also Basque has what is usually called an indefinite article. The marker, *bat*, originates from the numeral ‘one’ as well. Because its use with specific and nonspecific referents appears to be restricted by discourse-pragmatic factors, I treat *bat* as a presentational article. Importantly, the following examples will show that it occurs with both specific and nonspecific referents in those situations, in which the referential article is not sufficient to signal that the referent is non-identifiable.

Example (729) shows *bat* marking a specific referent. Even though there is a particular and salient referent evoked by the expression *gizon* ‘man’, the use of the presentation article marks it as non-identifiable.

(729) Badago [gizon bat] ate-an.
 is an ART:PRES door-ART:REF.LOC
 ‘There is a man at the door.’
 Basque (Hualde & Ortiz de Urbina 2003: 368)

Examples (730) and (731) show two contexts in which the use of *bat* instead of the referential article highlights that the identity of the referent is unknown, not only by the hearer but also to the speaker. Moreover, in (730), the identity of *peasant* ‘peasant’ is simply irrelevant besides its membership in a certain group of people. In (731), on the other hand, the use of *bat* highlights that also the speaker lacks relevant knowledge to properly identify the referent of *gizon* ‘man’

(730) [Baserritar-ren **bat**-ek] sortub-a izango da kontu ori ere.
 peasant-GEN ART:PRES-ERG create.PFV-ART:REF be.FUT AUX story that also
 ‘This story too has probably been made up by some peasant.’
 Basque (Hualde & Ortiz de Urbina 2003: 263)

(731) Izena ahantzi dudan [gizon **bat**-ek] erran dit.
 name forget AUX.COMP man ART:PRES-ERG tell AUX
 ‘A man (whose) name I have forgotten told me.’
 Basque (Hualde & Ortiz de Urbina 2003: 780)

In (732), the use of *bat* emphasizes that the referent of *izen* ‘name’ is not identifiable by the hearer, and that the speaker chooses not to disclose its identity.

(732) Ezagun baino ezagunago den [izen **bat**] aipa nezake.
 known than known.more is.that name ART:PRES quote AUX:POT
 ‘I could mention a name which is more than known.’
 Basque (Hualde & Ortiz de Urbina 2003: 276)

All previous contexts showed *bat* used to mark referents as specific. In examples (733) and (734), the referents in question, *etxe* ‘house’ and *errota* ‘wheel’, are ambiguous between a specific and a nonspecific interpretation. In (735), the referent of *kotxe berri* ‘new car’ is nonspecific. All three referents are marked by *bat* as well.

(733) Bilbon [etxe **bat**] erosi behar dut.
 Bilbao.LOC house ART:PRES buy need AUX
 ‘I need to buy a house in Bilbao.’
 Basque (Hualde & Ortiz de Urbina 2003: 694)

(734) Sarthu eta berehala ... ikusiko zinduen [errota **bat** gaitza].
 enter and immediately ... see.FUT AUX:PST wheel ART:PRES big
 ‘After entering, you would have found ... a big wheel.’
 Basque (Hualde & Ortiz de Urbina 2003: 267)

(735) [Kotxe berri **bat**] nahi dut.
 car new ART:PRES want AUX
 ‘I want a new car.’
 Basque (Hualde & Ortiz de Urbina 2003: 301)

In contrast to the Tepehua system presented in the previous section, Basque does not restrict the use of the presentational article to nominal expressions in the singular. Although called a quantifier in Hualde & Ortiz de Urbina (2003), Trask (2003: 122) explicitly notes that “[t]he quantifier *batzuk* ‘some, several’ [...] is formally a plural of this *bat*.” According to its distribution with

discourse-prominent specific and nonspecific referents, I treat *batzuk* as the plural form of the presentational article. Examples (736) to (740) illustrate the use of *batzuk* with specific referents whose precise identity is unknown or irrelevant, the referents themselves being relevant in the discourse situation.

- (736) Zergatik ez erregutu hari [bitxi **batzuk**] utz ziezazkion?
 why not beg 3SG.DAT jewel **ART:PRES.PL** lend AUX.SUBJ.COMP
 ‘Why not beg her to lend her some jewels?’
 Basque (Hualde & Ortiz de Urbina 2003: 503)
- (737) Gaur [gauza inportante **batzuk**] ahaztu dituzu.
 today thing important **ART:PRES.PL** forget AUX.S:2.O:3PL
 ‘Today you forgot important things.’
 Basque (Hualde & Ortiz de Urbina 2003: 576)
- (738) [Herri **batzuk**], 500 herritar baino gutiago dituzten-a-k, egoera
 village **ART:PRES.PL** 500 inhabitant than less have.COM-ART:REF-PL situation
 txarre-an dira.
 bad-LOC are
 ‘Some villages, which have less than 500 inhabitants, are in a bad situation.’
 Basque (Hualde & Ortiz de Urbina 2003: 805)
- (739) Asko aldiz entzun dukezu [gizon zoro **batzuen**] aho-tik solas hau.
 many time hear.PFV AUX:PRS.POT man crazy **ART:PRES.PL.GEN** mouth-ABL word this
 ‘You may have heard this word many times from some crazy people.’
 Basque (Hualde & Ortiz de Urbina 2003: 258)
- (740) [Ezezagun **batzuek**] lagundu gaituzte.
 unknown **ART:PRES.PL.ERG** help AUX.TR
 ‘Unknown people helped us.’
 Basque (Hualde & Ortiz de Urbina 2003: 413)

Returning to the bigger picture of the article system in Basque including a referential and a presentational article, several studies have examined the development of both articles (although necessarily as together as a system). For both articles, previous studies have suggested that they developed in contact with Spanish. The referential article can be traced back to the distal demonstrative which is *a* in Western and (*h*)*ura* in Central and Eastern Basque (Manterola 2007: 6). According to Lapesa (1960) and Epstein (1994), the referential article in Basque developed as a definite article around the same time when the definite article emerged in other neighboring Romance languages. While it is not relevant here whether or not the development of the referential article in its early, definite stage was contact-induced, it is important that, indeed, it started as a

definite article and extended its referential functions to mark specific and nonspecific referents in Modern Basque.

As for the development of the presentational article, it does not seem to be clear when and how exactly it developed. However, according to Manterola (2007: 9), both its morphology and diachronic sources attest that it is not a recent development. While its current plural form is *batzuk*, containing the singular *bat* and the plural marker *-k*, it used to have the form *batzu*. The latter form contains a collective suffix *-zu* whose productivity already started to decline in the Middle Ages according to Michelena (1971). In addition, the use of *batzu* as an indefinite marker in the plural is attested across various dialects of Basque from the beginning of its documentation in 1545 (Manterola 2007: 9). It is plausible that the use of the plural form *batzu* as an indefinite marker started out from a partitive meaning. The use of the singular numeral form *bat* appears to be a younger development. While there is no consensus on whether or not this is an independent or a contact-induced development, replicating the use of *un(os)/una(s)* in Spanish as suggested by Michelena (1971), its use in Modern Basque is more restricted than the use of *un/a* in Spanish (Trask 2003: 122).

The important point concerning systems with overlapping articles is that we see that a former definite article that has developed into functionally much broader referential article. In addition, we Basque has an old plural indefinite marker that has been extended to the singular and that is now used as a presentational article in Modern Basque. This means that we cannot be certain that the extension of the definite article to a referential one caused the development of an indefinite article in Basque, but this extension, as well as the presence of an indefinite article in Spanish and the availability of the partitive/indefinite plural marker *batzu* are three factors that certainly facilitated its development in those contexts in which the referential article was no longer sufficient to highlight the non-identifiability of the referent.

9.3.3 Maori

Maori is an Oceanic language spoken in New Zealand. According to the Endangered Languages Project, Maori has approximately 127,000 speakers.⁶ Especially its determiner system (and related questions about phrase structure) have been the focus of previous theoretical studies, most notably Chung & Ladusaw (2004), Chung et al. (1995), de Lacey (2001), Polinsky (1992), Waite (1994).

Traditionally, Maori has been described as having the definite article *te* and the indefinite article *teetahi*, with *he* as some kind of second indefinite marker. Even though in her reference grammar, Bauer (1993: 109-111) introduces the three markers with those labels, she notes that they are not entirely accurate:

⁶Cf. <http://endangeredlanguages.com/lang/3571>.

“*te* [is] usually described as ‘definite singular’, but also used as the generic article (with plural nouns in those few instances where the noun inflects for plural), and probably more accurately described as the “elsewhere” article—*te* is used when the more semantically specific articles are inappropriate.” (Bauer 1993: 109).

Similarly, Harlow (2012: 141) mentions that “*te* also appears in contexts where the speaker has in fact no specific entity in mind, and thus is indefinite.” As will be shown in this Section, *te* is indeed better described as an inclusive-specific article because it occurs in definite and specific contexts, while not being used with nonspecific referents.

The so-called indefinite article *tētahi* will be shown to correspond to a presentational article, being used only with certain types of specific or nonspecific referents. The last relevant determiner is *he*, cognate of the predicate marker *he* in Rapa Nui. In Maori, *he* is usually also called an indefinite marker or determiner (e.g. Bauer 1993: 110, Yamada 2014: 47), probably in lack of a better term. All studies note that *he* is restricted syntactically in a certain way and used to mark referents as nonspecific in opposition to *tētahi* in certain contexts (e.g. Bauer 1993: 109). The theoretical approaches to *he* can roughly be divided into two main camps. Some scholars have argued that *he* is a tense or aspect or predicate marker (Polinsky 1992, Waite 1994). Others have analysed *he* as a determiner; to be precise, often as a nonspecific determiner or article. (Chung & Ladusaw 2004, Chung et al. 1995, Harlow 2012, de Lacey 2001).

The most suitable analysis of its syntactic status is not relevant for the present study and will not be discussed here. In terms of its functions, however, I will follow Polinsky (1992) in arguing that *he* rather behaves like a predicate marker, albeit having extended its use to a number of other contexts that certainly make it a very atypical predicate marker. It may be tempting to analyse it as a nonspecific article, since it marks referents as nonspecific in certain contexts as will be shown below, but it also occurs with specific referents in other constructions. Therefore, one could argue that the nonspecific interpretation is the result of the entire construction including *he* rather than the semantic contribution of *he* itself. Treating *he* as a predicate marker is also only a shorthand, since it does not readily account for all of its uses, as was discussed in detail in Chung & Ladusaw (2004), Chung et al. (1995), de Lacey (2001). Besides the use of *he* as a predicate marker, its use in indefinite contexts, especially in relation to the article *tētahi* is conditioned by a number of syntactic, semantic, and probably also discourse-pragmatic factors. The semantic differences between the two markers are not very straightforward; in general, both can occur with specific as well as with nonspecific referents. Only in certain contexts, as will be shown below, *he* forces a nonspecific interpretation, while the presentational article *tētahi* only allows for a specific interpretation. Chung et al. (1995), offer a detailed account of the uses of *he* and call these uses “relational” as opposed to its “existential” use as a predicate marker. In relational

uses, *he* marks a referent within the scope of a question, negation, or modal operator, where the referent marked by *he* is necessarily interpreted as nonspecific.

In lack of a better label for such a hybrid marker between predicate marker and determiner, I will refer to *he* as a predicate marker for the purposes of the present study. What is important here is that it is not an article because its referential functions fall out from the constructions it is used in and from its interaction with the presentational article *tētahi* rather than from the semantic contribution of *he* itself. Before turning to the discussion of the two articles and *he*, Table 9.1 provides an overview of all the relevant forms.

Table 9.1: Maori determiners

	singular	plural
inclusive-specific article	<i>te</i>	<i>ngā</i>
presentational article	<i>tētahi</i>	<i>ētahi</i>
predicate marker	<i>he</i>	

9.3.3.1 The inclusive-specific article *te*

Examples (741) to (743) show the use of the inclusive-specific article *te* with definite referents. In (741b), *te* marks the anaphoric referent of *whare* ‘house’. Example (742) is an establishing context, introducing the referent of *kaimahi* ‘worker’ as identifiable based on the information provided in the same utterance. In (743), we see *te* occurring with a contextually unique referent in an equative context.⁷

- (741) a. Ko Wairangi te tangata whakamutunga ki te tomo i [te
TOP Wairangi ART:INSPEC man last to ART:INSPEC enter at ART:INSPEC
whare] ...
house ...
‘Wairangi was the last man to enter the house ...’
- b. Kaatahi ka rere a Wairangi ki te tute i [te whare], kore
then TA leap PROP Wairangi to ART:INSPEC shove do ART:INSPEC house NEG
rawa i ngaoko.
EMPH TA stir
‘Then Wairangi threw his weight against the side of the house, but it didn’t yield in
the slightest.’
(Bauer 1993: 153)

⁷The marker *ko* is glossed as a topic, equative, or identificational marker according to its function and the gloss in the source. I do not discuss these different functions of *ko* here, for details see Bauer (1993), Douglas (2018), Harlow (2012), de Lacey (2001), and references therein.

(742) Ka whakamiiharo anoo a Rewi ki [te kaimahi] whakahaere i te
 TA wonder.at again PROP Rewi to [ART:INSPEC worker] operate do ART:INSPEC
 miihini uta i ngaa raakau
 machine load do ART:INSPEC.PL tree
 ‘Rewi marvelled again at the worker operating the machine loading the trees.’
 Maori (Bauer 1993: 460)

(743) Nō te mutunga o 1899 ko ia [te minita mō ngā
 belong ART:INSPEC end of 1899 EQ 3SG ART:INSPEC minister belong ART:INSPEC.PL
 take Māori].
 affairs Māori.
 ‘From the end of 1899 he was the minister of Māori affairs.’
 Maori (Douglas 2018: 22)

The inclusive-specific article *te* occurs with specific referents as well, which is explicitly mentioned by Harlow (2012) and Bauer (1993). Examples (744) to (748) show the use of *te* to mark referents as specific indefinite.

(744) Ka takai-a, ka kawea, ka whaka-iri-a ki runga ki [te kauere].
 TA wrap-PASS PST carry-PASS TA CAUS-hang-PASS to top to ART:INSPEC puriri
 ‘They wrapped him up and took him and suspended him in a puriri tree.’
 Maori (Bauer 1993: 444)

(745) I tārai-a te waka ki [te toki pounamu].
 TA carve-PASS ART:INSPEC canoe PREP ART:INSPEC axe jade
 ‘The canoe was carved with a greenstone axe.’
 Maori (Harlow 2012: 157)

(746) ... ka tuku-a e te paa [te karere] ki roto o Hauraki.
 ... T/A send-PASS by ART:INSPEC pa ART:INSPEC messenger to inside GEN Hauraki
 ‘... the pa sent a messenger to Hauraki.’
 Maori (Bauer 1993: 354)

(747) Ka tango te tangata raa i [te koowhatu] hei kuru i a ia.
 T/A take ART:INSPEC man DIST ACC ART:INSPEC stone for pelt ACC PROP 3SG
 ‘That man grabbed a stone to pelt him with.’
 Maori (Bauer 1993: 354)

(748) Ka tango anoo taua kereruu i [te hua raakau].
 T/A take again ANA pigeon ACC ART:INSPEC fruit tree.
 ‘That pigeon again picked a berry.’
 Maori (Bauer 1993: 354)

Moreover, Chung et al. (1995: 433) note that *te* can be used to introduce new discourse-referents, which is probably the most prototypical context for specific indefinite expressions. Example (749) illustrates this.

- (749) Ka whānau [ngā tāngata] rā, ko [te tuakana], ko
 TA be.born ART:INSPEC.PL person there IDENT ART:INSPEC older.sibling IDENT
 [te teina].
 ART:INSPEC younger.sibling
 ‘Once there were born two men, an older and a younger brother.’
 Maori (Chung et al. 1995: 433)

9.3.3.2 The presentational article *tētahi*

In addition to the inclusive-specific article *te*, Maori uses the presentational article *tētahi* to mark specific and nonspecific referents that are also discourse-prominent. Examples (750) and (751) show *tētahi* with a specific referent. In (750), *tētahi* occurs in the prototypical context of introducing an important protagonist of the narrative.⁸ In example (752) *tētahi* occurs with the nonspecific referent of *raruraru kootahi* ‘one problem’.

- (750) Na, tērā [tētehi wahine puhi], ko Pare te ingoa, he tino
 now over.there ART:PRES woman virgin IDENT Pare ART:INSPEC name PRED very
 rangatira taua wahine.
 chiefly ANA woman
 ‘Now once there was a woman, called Pare, who was a puhi (virgin). This woman was of
 very high birth.’
 Maori (Chung & Ladusaw 2004: 32)
- (751) Ko too wahine kua hara ki [teetahi tangata ko Tupeteka te
 TOP 2SG.GEN woman TA sin INSTR ART:PRES man PREP Tupeteka ART:INSPEC
 ingoa].
 name
 ‘Your wife has sinned with a man called Tupeteka’
 Maori (Bauer 1993: 279)
- (752) Kaaore anoo [teetahi raruraru kootahi] kia paa.
 NEG again ART:PRES problem one SUBJ arise
 ‘Not a single problem had arisen.’
 Maori (Bauer 1993: 297)

⁸Note that *tētehi* is a dialectal variant of *tētahi* (Chung & Ladusaw 2004: 31).

9.3.3.3 *He* as a predicate marker

Examples (753) and (754) show the use of *he* to mark nominal predicates. In (755), *he* is used with a predicative adjective.

(753) [he tangata pai] a hoone.
PRED person good PROP John
'John is a good person.'
Maori (Polinsky 1992: 234)

(754) ko te raakau i moe ai ia [he raakau nui].
TOP ART:INSPEC tree TA sleep PRO 3SG PRED tree big
'The tree he was lying under was a great tree.'
Maori (Polinsky 1992: 233)

(755) [He reka] ngā kai.
PRED sweet ART:INSPEC.PL food
'The food is delicious.'
Maori (Harlow 2012: 143)

An additional argument by Polinsky (1992: 234) for the analysis of *he* as a predicate marker comes from appositions as in (756). She argues that *he taane* 'a boy' is in fact a secondary predicate in apposition to the primary predicate, i.e. a conjoined clause with a coreferential subject.

(756) ka whaanau taana tamaiti, [he taane].
TA be.born POSS:3SG child PRED male
'She gave birth to her child, a boy.'
Maori (Polinsky 1992: 234)

Another type of contexts suggesting that *he* is involved in marking predication rather than referentiality is its use following "typical predication markers" such as 'like', 'such as', and 'as' according to Polinsky (1992: 236) This is shown in examples (757) and (758), where *he* occurs with a noun following the preposition *me* 'as, like'. Note that the use of *he* parallels the use of cognate *he* in Rapa Nui in such contexts (cf. example (614) in Section 7.2.1).

(757) ka rongo i te reo, [me he reo tanata] nei.
TA hear ACC ART:INSPEC voice as PRED voice man DEM
'They hear voices, it was like human voices.'
Maori (Polinsky 1992: 237)

(758) [me he kawau puu] te aahua.
as PRED chief ART:INSPEC appearance
'[He] looked like a chief.'
Maori (Polinsky 1992: 237)

(759) [He mana tipua] o Māui.
 PRED power abnormal of Maui
 ‘Maui possessed *abnormal* powers.’
 Maori (Chung & Ladusaw 2004: 42)

Two more types of uses of *he* fall out from analysing it as a predicate marker. Example (760) shows that the use of *he* forces a generic interpretation. In this case, reference to individuals is excluded, and the concepts and their properties are evoked. This is compatible with the functions of *he* as a predicate marker, highlighting the properties of the concepts evoked.

(760) whatungarongaro [he tangata], toituu [he whenua].
 disappear PRED man permanent PRED land
 ‘Man disappears but land remains. / *Some man disappears but a land remains.’
 Maori (Polinsky 1992: 232)

Analysing *he* as a predicate marker can further account for its use that differs from the use of the article *te* in contexts like (761). The subject can be marked as specific by the inclusive-specific article *te* in the default position, i.e. in VSO order. In contrast, the use of *he* in this position is less felicitous. As can be seen in (761b), it becomes felicitous in a construction in which the subject is left-dislocated, preceding the verb. Polinsky (1992) proposes to analyse this structure as a predicate (*he tangata*) with a headless relative clause (*ka haere ki te moana*).⁹

(761) a. i haere [te / ?he tangata] ki te moana
 TA go ART:INSPEC / PRED man to ART:INSPEC ocean
 ‘A man went to the ocean.’
 b. [he tangata] ka haere ki te moana.
 PRED man TA go to ART:INSPEC ocean
 ‘He who went to the ocean was a man.’
 Maori (Polinsky 1992: 232, 234)

Analysing such constructions as involving headless relative clauses could also account for the use of *he* in certain nonspecific contexts. Consider example (762). The sentence in (762a) shows an affirmative utterance with a definite referent marked by the inclusive-specific article *te*. In (762b), we see that *te* has wide scope over the negation. Using *he* in the negated construction as in (762c), however, necessarily leads to a nonspecific interpretation of the referents of *tangata* ‘man’. A more faithful translation of the utterance in (762c) would thus be ‘[Those who] are men are not at work’.

⁹The construction in (761b) is syntactically similar to a topic construction, in which the subject is left-dislocated and marked by *ko*, the topic marker. The closely related language Rapa Nui has the same topicalization construction (Kieviet 2017a: 407), which also uses *ko* as a topic marker. Moreover, the predicate marker *he* in Rapa Nui can be used instead of *ko* in such topic constructions as well. Interestingly, Kieviet (2017a: 410-411) notes that Rapa Nui *he* is used in such constructions with generic statements or in cases in which the referent is not identifiable.

- (762) a. i te mahi [ngaa taangata].
 LOC ART:INSPEC work ART:INSPEC.PL man
 ‘The men are at work.’
- b. kaahore [ngaa taangata] i te mahi.
 NEG ART:INSPEC.PL man LOC ART:INSPEC work
 ‘The men aren’t at work.’
- c. kaahore [he taangata] i te mahi.
 NEG PRED man LOC ART:INSPEC work
 ‘Nobody is at work. / *Some men are not at work.’
 Maori (Polinsky 1992: 237)

9.3.3.4 *Tētahi* and *he* with specific referents

The predicate marker *he* can also occur with specific referents. While the fact that *he* only occurs with subjects of intransitive or passive verbs had been known, Chung et al. (1995) draw a much more variegated picture of the distribution of *he* in what they call its “existential” uses. Existential *he* can only occur with nominal expressions that are syntactically subjects which do not correspond to the agent argument. The most important types of contexts in which the subject is a non-agent are: passives of transitive verbs, unaccusative intransitive verbs, psychological verbs, adjectives, and stage-level predicates such as ‘be cold’, ‘be angry’ (Chung et al. 1995: 438-441, 445). Thus, *he* cannot occur in a position following a preposition, it cannot occur in an object position or with a subject of a transitive verb. Such restrictions do not apply to *tētahi*, and the presentational article is often used in those cases in which the use of the predicate marker *he* is blocked syntactically.

As is suggested by the label of “existential”, we find *he* with specific referents in existential constructions. This is shown in (763).

- (763) [He tuna] no roto i nga awa, ā, [he manu] no runga i
 PRED eel TA.of inside ACC ART:INSPEC.PL river and PRED bird TA.of top ACC
 nga maunga.
 ART:INSPEC.PL mountain
 ‘There were eels in the rivers and birds in the ranges.’
 Maori (Chung & Ladusaw 2004: 42)

Example (764) shows another utterance with *he* used to mark a referent as specific as part of an episodic event.

- (764) i kai-nga e hoone [he kai].
 TA eat-PASS AG John REF food
 ‘John ate some food.’
 Maori (Polinsky 1992: 231)

The two contexts presented in (765) and (766) show a similar use of both *he* and the article *tētahi*. In those two cases, there is no difference between the semantic contributions of the two markers.

- (765) a. Kua tae mai [he manuhiri] ki taku kāinga.
 TA arrive to.here PRED guest to my house
 ‘Some visitors came to my house.’
- b. Ka tae mai [tētahi taraka] tino nui.
 TA arrive to.here ART:PRES truck very big
 ‘A huge truck came here.’
 Maori (Chung & Ladusaw 2004: 21)

- (766) a. Kua riro [he pukapuka a Mere].
 TA be.taken PRED book of Mere
 ‘A book of Mere’s was taken.’
- b. Kua riro [tētahi pukapuka a Mere].
 T be.taken ART:PRES book of Mere
 ‘A book of Mere’s was taken.’
 Maori (Chung & Ladusaw 2004: 27)

Example (767) shows that *he* can also be used to introduce protagonists of a narrative. Example (767) is the counterpart of examples (749) and (751) above, which showed the inclusive specific and presentational articles *te* and *tētahi* in a similar function.

- (767) Tērā [he tangata], ko Rua-rangi te ingoa, ko Tawhaitū
 over.there PRED person IDENT Rua-rangi ART:INSPEC name IDENT Tawhaitū
 te ingoa o tana hoa.
 ART:INSPEC name of his friend
 ‘Once there was a man called Rua-rangi; Tawhaitū was the name of his wife.’
 Maori (Chung & Ladusaw 2004: 32)

9.3.3.5 *Tētahi* and *he* with nonspecific referents

Both the presentational article *tētahi* and the predicate marker *he* occur with nonspecific referents. Regarding the latter, Chung et al. (1995) call such uses relational uses, as they always fall within the scope of negation, a question, or a modal operator, etc. Importantly, *he* is less restricted syntactically in such contexts. While it still only occurs with nominal expressions that are subjects, the semantic role is no longer important, meaning that relational *he* can also occur with subjects that are agents.

Examples (768) and (769) show the use of *tētahi* and *he* in imperative contexts, where the indefinite referents are necessarily interpreted as nonspecific. Analysing the construction involving *he* as a headless relative clause would translate into ‘Give me [what] is water.’

(768) Kawe-a atu [**tētahi** wai] mo-ku.
bring-PASS away **ART:PRES** water T.of-me
'Fetch me some water.'

Maori (Chung & Ladusaw 2004: 22)

(769) Homai [**he** wai] mo-ku.
give.PASS **PRED** water T.of-me
'Give me some water.'

Examples (770) and (771) show two hypothetical conditional contexts. Chung & Ladusaw (2004: 34) explicitly state that "the indefinites in all these examples have narrow scope". In these two examples, we see the different syntactic distributions of *he* and *tētahi* at play. In (770) *he* occurs with the subject *tohora* 'whale' of the passive verb *kitea* 'see.PASS'. However, the referent of *kaititiro* 'lookout' is the agent expressed as an oblique argument and is marked by the presentational article *tētahi* instead. A similar situation is shown in (771), where the referents of *Māori* (subject) and of *mātauranga* 'education' (object) occur with *he* and *tētahi*, respectively.¹⁰

(770) Ka kite-a [**he** tohora] e [**tētahi** kaititiro], ka whakatū-ria te haki o
TA see-PASS **PRED** whale by **ART:PRES** lookout TA raise-PASS **ART:INSPEC** flag of
tōna kāinga.
his house
'If a lookout spotted a whale, he would raise the flag of his house.'

Maori (Chung & Ladusaw 2004: 34)

(771) Ki te whakawhiti [**he** Māori] ki tāwāhi ki te kimi i [**tētahi** mātauranga]
if cross.over **PRED** Maori to overseas **NONFIN** seek **ACC** **ART:PRES** education
mō-na, ... ko Hoani tonu te mea tuatahi ki te whakatū komiti.
TA.of-him ... **IDENT** Hoani still **ART:INSPEC** person first **NONFIN** set.up committee
'If a Maori went abroad to seek an education for himself, Hoani was always the first person
to set up a committee.'

Maori (Chung & Ladusaw 2004: 35)

The following pair of examples, (772) and (773), shows two contexts of a nonspecific referent within the scope of a question. Again, we find that both the presentational article *tētahi* in (772) as well as the predicate marker *he* in (773) are able to mark the referents in question.

(772) I kite anō koe i [**tētahi** mea rerekē] i to haerenga mai?
TA see indeed you **ACC** **ART:PRES** thing strange at your going to.here
'Did you *observe* anything strange when you came?'

Maori (Chung & Ladusaw 2004: 39)

¹⁰The complex expression *ki te* can be used in the sense of *if* but also to mark dependent nonfinite clauses, reflected by the glossing as either 'if' or 'INF' in example (771).

- (773) Kua rongo anō koe [he kōrero mo te aituā]?
 TA hear indeed you PRED talk TA.of ART:INSPEC accident
 ‘Have you heard anything *concerning* the accident?’
 Maori (Chung & Ladusaw 2004: 39)

To sum up, the examples of the last two sections showed that *tētahi* and *he* can be used in comparable semantic contexts, including the marking of specific as well as nonspecific referents. After presenting the semantic differences that do arise in certain types of contexts, we will return to potential explanations of the distribution of *tētahi* and *he* in those contexts where their use is not conditioned by semantic or syntactic factors.

9.3.3.6 The factors conditioning the use of *te*, *tētahi*, and *he*

In addition to the syntactic restrictions of *he*, which readily account for the use of *tētahi* in certain syntactic positions, there is a semantic difference between the two markers in certain other contexts.

As was shown in the previous sections, both the presentational article *tētahi* as well as the predicate marker *he* are used in affirmative existential constructions (cf. examples (750) and (753)). However, in the presence of negation resulting in the interpretation of the referent as nonspecific, only *he* can be used. This is shown in examples (774) and (775) for singular and plural expressions.

- (774) Kāore [he / *tētahi take kotahi].
 TA.not PRED / ART:PRES reason one
 ‘There’s no reason at all.’
 Maori (Chung & Ladusaw 2004: 43-44)

- (775) Kāhore [he / *ētahi taniwha].
 TA.not PRED / ART:PRES.PL taniwha
 ‘There are no taniwhas.’
 Maori (Chung & Ladusaw (2004: 44) based on Bauer (1993: 78))

The presentational article *tētahi*, on the other hand, marks an indefinite referent as specific within the scope of negation. This can be seen in (776) and (777).

- (776) Kāore ia i kite i [tētahi hō] e takoto ana i roto i ngā karaehe.
 not he TA see ACC ART:PRES hoe TA lie PRT at inside at ART:DEF.PL grass
 ‘He didn’t see a hoe (=there was a hoe that he didn’t see) lying in the grass.’
 Maori (Chung & Ladusaw 2004: 40)

- (777) Kaore [ētahi momo rākau] e tipu i konei.
 not ART:PRES.PL kind tree TA grow at here
 ‘*Certain* trees will not grow here.’
 Maori (Chung & Ladusaw 2004: 40)

Besides the syntactic differences and the different behaviour in the scope of negation, Chung et al. (1995) note a difference between the use of the presentational article *tētahi* and the predicate marker *he* resulting from discourse-pragmatic factors. They formulate the following discourse-based restriction for *he*, in contrast to *tētahi*:

“Given that *he*-indefinites have an existential use, it is striking (from the Indo-European perspective) that they are rarely the NPs chosen to introduce the first discourse referent of a narrative. The types of NPs that typically serve this function are instead specified by the definite article, a demonstrative, or the indefinite article *tētahi*. [...] A *he*-indefinite cannot introduce a referential argument that is in the spotlight; that is, a central focus of narration.” (Chung et al. 1995: 433-434)

Thus, they suggest that specific referents tend to be marked by *he* if they are not in the center of attention in the following discourse segment. When introducing an important protagonist in a narrative, the presentational article *tētahi*, the inclusive-specific article *te*, or a demonstrative are used instead. That the inclusive-specific article and the presentational article can introduce new discourse referents was shown in examples (749) and (750) above. The use of the demonstrative *tēnei* is reminiscent of establishing uses of *this* in English; an example is shown in (778).

(778) Ko [tēnei taniwha] nō Ngāti Whātua ki Kaipara.
 TOP this monster TA.of Ngāti Whātua at Kaipara
 ‘This taniwha was associated with Ngāti Whātua at Kaipara.’
 Maori (Chung et al. 1995: 433)

Chung et al. (1995) provide examples of the use of *he* marking specific referents that are not very discourse-prominent. They introduce the example shown in (779) as follows: “The story begins when Parekawa, a tohunga’s daughter, violates tapu by preparing food after having cut her father’s hair. Some visitors arrive when she is alone, and she is so embarrassed she cooks for them. She then tells her father:” (Chung et al. 1995: 433)

(779) Kua tae mai [he manuhiri] ki taku kāinga, kāore he tangata hei tahu kai.
 TA arrive to.here PRED guest to my house not PRED people so.as.to cook food
 ‘Some visitors had arrived at my house and there was no one to cook food.’
 Maori (Chung et al. 1995: 433)

Thus, the exact identity or nature of the referent of *manuhiri* ‘visitors’, marked by *he* is irrelevant in the story. Only the event of the visit has an important consequence, because it leads to Parekawa having to prepare food, even though this is against the rules. Example (780) shows the utterance that follows (779) in the narrative.

- (780) Ka tonoa mai e ia [he kai] mā-na. Ko aua kai he korokoro, he
 TA send-PASS to.here by him PRED food TA.of-her TOP ANA.PL food PRED lamprey PRED
 kōkopu.
 kōkopu
 ‘[The father felt compassion for his daughter, and] he sent some food for her. That food
 was lamprey and kōkopu.’
 Maori (Chung et al. 1995: 434)

In examples (780) and (781), the expressions *kai* ‘food’ and *tohunga* ‘tohunga’ are marked by *he*, even though they correspond to the attentional center of the following utterance. Chung et al. (1995: 434) note that after being referred to once or twice again in the immediately following utterances, those referents are not mentioned again in the narratives. The referent of *karere* ‘messenger’ in (782) is only mentioned this one time in the narrative; clearly, it is the function of the referent as a messenger rather than its exact identity what is relevant in the narrative.

- (781) Kātahi ka tiki-na [he tohunga] hei karakia. Ka tae mai te
 then TA fetch-PASS PRED skilled.person so.as.to chant TA arrive to.here ART:INSPEC
 tohunga, kātahi ka karakia-tia.
 skilled.person then TA chant-PASS
 ‘Then they fetched a tohunga to perform rituals over him. The tohunga came and per-
 formed the rituals.’
 Maori (Chung et al. 1995: 434)

- (782) Nō te oti-nga ka tonoa e rātau [he karere] ki te taniwha
 TA.of ART:INSPEC done-NMLZ TA send-PASS by them PRED messenger to ART:INSPEC
 rā ki te kī atu kia haere mai ki tana wahine ...
 monster there INF say away TA go to.here to his woman ...
 ‘When [the house] was completed they sent a messenger to the taniwha to tell him to come
 join his wife ...’
 Maori (Chung et al. 1995: 434)

Chung et al. (1995: 545) mention example (783) as a potential counter-example to *he* being used with specific referents of low discourse-prominence. This is because in (783) *he* occurs in the first utterance of a narrative, introducing its main protagonist. However, as was shown above, *he* functions as a predicate marker and is typically used in existential constructions (cf. examples (763), (767), (774), (775)). Thus, its use in cases like (783) could be attributed to the use of *he* as a predicate marker in existential constructions, where it marks the nominal expression *tangata* ‘person’ as a predicate rather than coding it as a specific referent. After all, the use of existential constructions to introduce new discourse referents, especially at the beginning of a narrative, is very common.

- (783) [He tangata] e haere ki te kāinga o Te Ruahine-mata-māori.
 PRED person TA go to ART:INSPEC house of Te Ruahine-mata-māori
 ‘There was a man who went to the house of Te Ruahine-mata-māori.’
 Maori (Chung et al. 1995: 454)

Summing up the important patterns, we saw that *he* only occurs with subjects (and only non-agent subjects with specific referents) and that the presentational article *tētahi* is used whenever *he* is syntactically blocked. Besides this and the difference in their referential functions in the context of negation, the last examples showed that *he* rather marks non-identifiable referents that are not very prominent. Often, it is the type of referent or its function within a given context that is relevant for the discourse. This discourse-pragmatic restriction could be the result of *he* being a predicate marker that is primarily used to refer to kinds and properties instead of individual referents. In contrast to *he*, the presentational article *tētahi* is used to mark referents that are non-identifiable but discourse-prominent, as its counterparts in Tepehua and Basque. In those cases, the identity of the referent is relevant, but the presentational article signals to the hearer that they do not yet know the identity of the referent.

Finally, we saw in the beginning of this section that the inclusive-specific article *te* can also be used to mark specific referents. Having focused mostly on the similarities and differences of *tētahi* and *he*, the use of *te* other than in definite contexts has received little attention in the literature. An account of the use of *te* is therefore somewhat speculative. However, an interesting trend emerges from examples (744), (745), (747), and (748), showing *te* with specific referents. The identity of those referents marked by *te* does not appear to be very relevant for the discourse in any of these examples. Moreover, in all four examples, *te* occurs with nominal expressions following a preposition (*ki* ‘with’) or the accusative marker *i*. Those are exactly the types of syntactic contexts in which *he* cannot be used. The alternative, the presentational article *tētahi*, may be too foregrounding for marking specific referents that are not prominent. Because Maori does not allow for bare nouns in argument positions, the inclusive-specific article could be used as some sort of fall-back determiner in such cases, where no other option is available and where the identifiability status of the referent is not very relevant either. In that sense, the description by Bauer (1993: 109) as an “elsewhere article”, at least with specific referents, may be a very appropriate description of *te*.

9.4 A typology of article systems

Similarly to the typology of articles presented in Section section 8.5, we can distinguish various article systems on the basis of how they cover the referential functions on the referential scale within a given language. This section will provide an overview of the attested systems in the

sample (and in certain cases, outside of the sample). There are 22 systems that need to be distinguished. For better clarity I divide them into four groups: (i) those that mark only functions in the definite domain, (ii) in the the indefinite domain, (iii) in both domains, and (iv) systems that have functionally overlapping articles (cf. Sections 9.2 and 9.3). I will also mention certain types of unattested article systems, restricting the discussion to the theoretically interesting systems.

9.4.1 Article systems marking the definite domain

The article systems that mark functions from the definite domain only are shown in Table 9.2. For each type, one language is given as an example (mostly one of the languages discussed in more detail in the previous chapters). The number in the rightmost column corresponds to the number of languages with that system in the sample.

Table 9.2: Article systems only marking the definite domain

	pragmatic definite			semantic definite		indefinite		example	N
	R _{anaph}	R _{recog}	R _{establ}	R _{sit.unique}	R _{cont.unique}	R _{spec}	R _{nspec}		
1		RECOG						Gooniyandi	3
2		ANA						Limbum	15
3			DEF					Mokpe	26
4		ANA			DEF _{WEAK}			unattested	
5					DEF _{WEAK}			unattested	

The first system consists of a single recognitional article. The three languages in the sample with such a system are all Australian languages. Section 8.2 showed that we find an very strong trend in Australian languages against articles, whereas recognitional articles appear to be an areal feature of Australia and Papunesia. The other two systems with a single article in Table 9.2 consist of a single anaphoric (2) or definite (3) article. These two systems appear to be fairly common in the world’s languages. Systems of type 1, 2, and 3 were presented in detail in Chapter 5.

System 4, consisting of an anaphoric article and a weak definite article is listed as unattested in Table 9.2. As was shown in Section 5.3, most of the languages discussed as having weak definite articles have no weak definite article in the strict sense. German and Fering have an additional indefinite article. The only potential candidate of a type 4 system is Icelandic, which does not use articles for the indefinite domain. However, the distinction between the anaphoric and the weak definite article in Icelandic is restricted to contexts with an evaluative adjective (cf. Section 9.1.1). In all other contexts, the definite article encodes both pragmatic and semantic definite referents, which is why I treat it as a different system of type 27 mentioned below. Although systems of

type 4 do not seem to be attested, there is no reason to categorically exclude their existence. For all we know, their development is very unlikely but not impossible, since the development of an anaphoric and a weak definite article does not depend on articles in the indefinite domain.

Systems with a single weak definite article (5) are equally marked as unattested in Table 9.2. In this case, however, there are reasons to exclude the existence of such an article system. Weak definite articles being extremely rare in the first place, all evidence points towards their emergence being dependent on the presence of another anaphoric marker that blocks the use of the former definite article in pragmatic definite contexts. This appears to be the only process that may lead to the development of a weak definite article (cf. Universal 6), which do not even develop in most cases in such a setting.

9.4.2 Article systems marking the indefinite domain

Table 9.3 shows the article systems that overtly code referential functions from the indefinite domain.

Table 9.3: Article systems only marking the indefinite domain

	pragmatic definite			semantic definite		indefinite		example	N
	R _{anaph}	R _{recog}	R _{establ}	R _{sit.unique}	R _{cont.unique}	R _{spec}	R _{nspec}		
6						EXSPEC		Angolar	3
7							INDEF	SL Malay	21
8						EXSPEC	NSPEC	Q'anjob'al	0
9							NSPEC	unattested	

Having a single exclusive-specific article (6) is attested, but only 3 languages in the sample have such a system. Angolar, a Creole language spoken in Sao Tome and Principe, is one of the languages in the sample with a system of type 6. The only article in the language is *uⁿa*. It is used to mark specific referents as in (784), while both definite (785) and nonspecific referents (786) are expressed by a bare noun.

(784) N ka tango [uⁿa θoya].
 1SG FUT tell ART:EXSPEC story
 'I will tell a story.'
 Angolar (Lorenzino 1998: 133)

(785) Nda ma vutuka fo mionga vuča kai.
 when again leave sea go.back house
 'When I came out of the sea again, I went back home.'
 Angolar (Lorenzino 1998: 168)

(786) Ši n ka vuna kiki e ngai n na ta ma n ka paθa wa.
if 1SG IRR catch fish big 1SG not know REL 1SG IRR happen not
'If I caught a big fish, I don't know what would happen to me.'

Angolar (Lorenzino 1998: 170)

It is much more common to have a single indefinite article (7), covering both specific and nonspecific referents. Out of 104 languages in the sample, 21 languages display this system, making it one of the most frequent articles systems in the world's languages. An example of this type of system is Sri Lanka Malay, which was described in detail in Section 6.1.2.

The crosslinguistic evidence for system 8, consisting of an exclusive-specific and a nonspecific article, is rather weak and listing it here as attested is somewhat tentative. Q'anjob'al appears to have an indefinite article that developed into an exclusive-specific article as a result of being blocked in nonspecific contexts, where a new more complex nonspecific article seems to have emerged. If so, then Q'anjob'al has an article system of type 8 because it does not use any additional article in the definite domain. However, the picture is somewhat more complicated, since Q'anjob'al also has both sortal and numeral classifiers. The sortal classifiers distinguish natural gender and social status for human referents, and they classify other types of concrete referents according to physical properties (Mateo Toledo 2017: 545). Most importantly for the purposes of this section, sortal classifiers in Q'anjob'al do not indicate the referential status of the noun per se; however, nouns can also occur without a classifier (also in the absence of any other pronominal determiner). It may thus well be that their use is in some way conditioned by discourse-pragmatic factors, at least in part. Nevertheless, the existence of system 8 should not be categorically excluded but expected to be very rare crosslinguistically, simply because nonspecific articles are very rare to begin with.

System 9, consisting of a single nonspecific article, is unattested. In this case, there is a theoretical motivation to exclude the existence of this type of systems. As was argued in Section 9.1.2, nonspecific articles usually originate from indefinite articles whose use is later restricted to nonspecific contexts. Also, there is no evidence for other sources of nonspecific articles. This observation can be formulated as in Universal 6.

Universal 5 *Nonspecific articles depend on specific articles*

Nonspecific articles only occur in article systems with either inclusive-specific or exclusive-specific articles.

9.4.3 Article systems marking the definite and indefinite domain

Table 9.4 shows the article systems that overtly code referential functions from both the definite and the indefinite domain without covering the entire referential space. In other words, some referential functions are left unmarked in these systems.

Table 9.4: Article systems marking the definite and indefinite domain I

	pragmatic definite			semantic definite		indefinite		example	N
	R _{anaph}	R _{recog}	R _{establ}	R _{sit.unique}	R _{cont.unique}	R _{spec}	R _{nspec}		
10	ANA					EXSPEC		Akan	2
11	ANA						INDEF	Teiwa	2
12				INSPEC				Bemba	0
13				DEF		EXSPEC		Agta	5
14				DEF _{WEAK}		EXSPEC		unattested	
15				DEF _{WEAK}		INDEF		unattested	
16				INSPEC _{WEAK}				unattested	
17				REF _{WEAK}				unattested	

Systems 10 and 11 have an anaphoric article and an exclusive-specific or indefinite article. Therefore, semantic definite referents as well as nonspecific referents in the case of system 10 remain unmarked. Those systems appear to be rather rare crosslinguistically, with two examples each in the sample. One example of system 10 is Akan, discussed in detail in Sections 5.2.3 and 6.3.2.

Teiwa (Timor-Alor-Pantar, Indonesia) is a language with an article system of type 11, i.e. with an anaphoric article and an indefinite (presentational) article. Examples (787) to (790) show how the different referential functions are expressed in such a system. In (787), we see the use of the anaphoric article *waal* with an anaphoric referent. However, this article is not used with semantic definite referents, as is shown in (788). In the indefinite domain, Teiwa makes use of the presentational article *nuk*. Example (789) shows *nuk* with a specific referent, example (790) with a nonspecific one.

(787) a. ...ana' maan si ki uwaad nuk yaa, **bif** ga'an tu'uk.
 ...long.time NEG SIM eagle big ART:PRES descend child 3SG knock
 'Not long [after that] a big eagle comes down and picks the child.'

b. [Bif waal] ta ba'-an yaa, ...
 child ART:ANA TOP fall-REAL descend ...
 'That child falls down, ...'

Teiwa (Klamer 2010: 433)

- (788) Iman yix-ei yaqai yir g-or an ma gi.
 they descend-REAL down.below water 3S-tail **market** come go
 ‘They went down to the market at the mouth of the river.’
 Teiwa (Klamer 2010: 436)
- (789) Ta rus waal bif un ga-bir-an gi ma [saf nuk] wan moxod.
 TOP deer ART:ANA child CONT 3SG-run-REAL go come river.bank ART:PRES be drop
 ‘So that deer takes the child to a river bank and drops it.’
 Teiwa (Klamer 2010: 433)
- (790) Naree maan, xoran si ma ha gi [ya’ siis nuk]
 grandfather NEG thus SIM come 2SG go kind.of.small.bamboo dry ART:PRES
 ga-uyan pin aria’, ...
 3SG-look.for hold arrive ...
 ‘Grandfather no, if so you go look for dry bamboo to bring here ...’
 Teiwa (Klamer 2010: 340)

Systems 12 and 13 both overtly code pragmatic and semantic definites as well as specific referents, leaving only nonspecific referents unmarked by articles. System 12 consists of a single inclusive-specific article, encoding all definite and specific functions, while system 13 combines a definite article with an exclusive-specific article. Both types of systems not being very common across languages, the latter type appears to be somewhat more frequent than the former. In the sample, no language has an article system of type 12. The only example that I am aware of is Bemba (cf. Section 7.1.1). This is partly due to the fact that inclusive-specific articles are crosslinguistically very rare to begin with (cf. Section 8.1). Usually, languages with an inclusive-specific article also have a nonspecific article.

An article system of type 13 was not presented in the previous chapters. One example from the sample is (Dupaningan) Agta, an Austronesian language spoken in the Philippines. Agta has a definite article *i*. Examples (791) and (792) show *i* with an anaphoric and a contextually unique referent, respectively. In (793), we see the exclusive-specific article =*a*, which is not used to encode nonspecific referents. The absence of articles with the latter can be seen in (794).

- (791) a. katandi=an=mi a mag-sikaw=bi=dan ihe, Bulos=aye.
 know=LOC=1PL.EX.GEN LNK ACT-swidden=also=already here Bolos=this
 ‘We also understand how to swidden farm here, here in Bolos.’
- b. kalan-an=tam=bi=dan hito [i mag-sikaw].
 copy-LOC=1PL.IN.GEN=also=already there ART:DEF ACT-swidden
 ‘We can all copy the swidden farming.’
 (Robinson 2008: 324)

(792) ballay-an=di i 'anak hidi=a [i kemukasan].
 await-LOC=3PL.GEN ART:DEF children PL=SPC ART:DEF new.year
 'The children are looking forward to the new year.'

(Robinson 2008: 85)

(793) mag-hen=kano ha rukab mangilen=a
 AV-live=HSY OBL cave ghost=ART:EXSPEC
 'They say a ghost was living in a small cave'

(Robinson 2008: 94)

(794) awan=bi a nag-ga-galgal i ulitaw=a ha ogsa
 NEG=also LNK COMPL.AV-MULT-slice ART:DEF young.man=ART:EXSPEC OBL deer
 'The young man has never butchered a deer.'

(Robinson 2008: 96)

Besides other logically possible combinations of articles along the referentiality scale to cover parts of both the definite and the indefinite domain not shown in Table 9.4, we can distinguish 4 more relevant system (14-17) that are not attested. They are all unattested due to the dependency of weak articles on the presence of an anaphoric article. Thus, we do not find a weak definite article together with an exclusive-specific (14) or an indefinite (15) article in the absence of an anaphoric article. Neither are systems of a single weak inclusive-specific article (16), nor a single weak referential article (17) attested. Note that such weak articles are crosslinguistically extremely rare in the first place, with only a handful of attested weak definite articles, one single weak inclusive-specific article, and no attested case of a weak referential article (whose existence should not be categorically excluded but only expected to be extremely rare). Based on the patterns uncovered for weak definite and weak inclusive-specific articles, we can however predict that if a language has a weak referential article, the article system will also include an anaphoric article. This allow us to formulate Universal 6:

Universal 6 *Weak articles depend on anaphoric articles*

Weak definite, weak inclusive-specific, and weak referential articles only occur in article systems that include an anaphoric article as well.

Table 9.5 shows the article systems that overtly mark all referential functions on the scale.

Table 9.5: Article systems marking the definite and indefinite domain II

	pragmatic definite			semantic definite		indefinite		example	N
	R_{anaph}	R_{recog}	R_{establ}	$R_{\text{sit.unique}}$	$R_{\text{cont.unique}}$	R_{spec}	R_{nspec}		
18	REF							Rapanui	6
19	INSPEC					NSPEC		Ayoreo	2
20	DEF					INDEF		Kaqchikel	9
21	DEF					EXPEC	NSPEC	Lakota	2
22	ANA	DEF _{WEAK}			INDEF		Fering	0	
23	ANA	INSPEC _{WEAK}				NSPEC		Tongan	1
24	ANA	REF _{WEAK}						unattested	
25	ANA	DEF _{WEAK}			EXSPEC	NSPEC		unattested	

System 18 consists of a single referential article that encodes all referential functions on the scale; such a system is well attested and the second most frequent type of systems covering the entire referential space in the sample. This study presented five examples of referential articles in Rapa Nui, Baure, Halkomelem, Huehuetla Tepehua, and Basque (cf. Sections 7.2 and 9.3). System 19, a combination of an inclusive-specific and a nonspecific article is very rare across languages but attested in two languages in the sample, namely Siar Lak and Ayoreo. The Ayoreo system was described in detail in Sections 6.4.1 and 7.1.2. The combination of a definite and an indefinite article (20), on the other hand, is more common in the world’s languages; it corresponds to the most frequent of such complex systems that cover the entire referential space. System 21 uses 3 distinct articles to express definite, specific, and a nonspecific functions. This type of system is very rare as well, with two examples in the sample. Data from both languages was presented in the previous chapters (cf. Section 6.3.1) for Biak and Sections 5.3.3 and 6.4.2 for Lakota).

The last two attested systems of this series (22 and 23) both involve an anaphoric article together with a weak article. Besides the anaphoric article, systems of type 22 have a weak definite and an indefinite article. Fering is an example of such an article system (cf. Section 5.3.1), which is very rare crosslinguistically and not attested in the sample. Type 23 has an additional weak inclusive-specific article together with a nonspecific article. Section 7.1.3 showed that Tongan is likely to have such a system, which seems crosslinguistically even rarer than system 22.

The last two systems in Table 9.5, 24 and 25, also contain weak articles; they are even rarer than the previous two systems in that they are not attested to the best of my knowledge, but their existence should not be excluded categorically. They involve a combination of many different developments which are rare on their own, making their co-occurrence even rarer. System 24 consists of an anaphoric article and a weak referential article. Such a system would develop

from an article system with a referential article and an anaphoric marker whose use becomes conventionalized with pragmatic definites, blocking the use of the referential article in those contexts. While this article system as such is not attested, similarly to the situations described for Lakota and Hausa, Rapa Nui uses anaphoric and other demonstratives in certain cases instead of the referential article especially with pragmatically definite referents to highlight their discourse-prominence when the referential article is not sufficient to do so (cf. [Kieviet 2017b](#)). The last system in Table 9.5, system 25, consists of an anaphoric, a weak definite, an exclusive-specific, and a nonspecific article. This system is not attested either. Again, its status as unattested is rather due to the fact that different already very rare developments must have taken place to lead to such an article system, whose existence should nevertheless not be excluded categorically.

9.4.4 Article systems with functional overlap

The last series of article systems distinguished here is shown in Table 9.6. These systems all include articles with overlapping referential functions. There are many other logically possible combinations of articles with overlapping functions. For this series, I only list the attested ones. In general, the figures in Table 9.6 show that such article systems are not very common in the world’s languages.

Table 9.6: Article systems with functional overlap

	pragmatic definite			semantic definite		indefinite		example	N
	R_{anaph}	R_{recog}	R_{establ}	$R_{\text{sit.unique}}$	$R_{\text{cont.unique}}$	R_{spec}	R_{nspec}		
26			DEF					Urama	1
		ANA							
27			DEF					Icelandic	0
		ANA		DEF _{WEAK}					
28			DEF					German	1
		ANA		DEF _{WEAK}		INDEF			
29			DEF			INDEF		Lavukaleve	2
		RECOG							
30				INSPEC				Maori	1
						INDEF			
31				REF				Basque	2
						INDEF			

System 26, consisting of a definite and an anaphoric article, was described in Section 5.3.4 with data from Urama. The two systems 27 and 28 involve a weak definite article. These two cases

of overlap differ from the other ones shown here in that they do not involve distinct markers that overlap in their functions. Rather, the weak definite article in Icelandic is only restricted to semantic definites under certain circumstances and functions as a definite article otherwise. In the German system, the distinction between an anaphoric and a weak definite article is not always available either, making the strong definite article (i.e. the article in the non-contracted combination of preposition and article) a general definite article in all other contexts. Also the functional overlap of the definite and recognitional article in system 29 tentative to a certain extent. More research will be required to determine if those articles really overlap in their functions or if the definite article in systems of type 29 simply does not encode recognitional referents.

The last two systems, system 30 and 31, involve a functional overlap in the coding of specific referents, as was described in detail in Section 9.3 for Tepehua, Basque, and Maori. They consist of an inclusive-specific (30) or a referential (31) article used together with an indefinite (or presentational) article. Similarly to the situation with weak articles mentioned above, the functionally broader inclusive-specific or referential articles are replaced by the presentational article in order to highlight a referent as non-identifiable in certain cases.

9.4.5 The crosslinguistic distribution of article systems

This section will take a closer look at the crosslinguistic distribution of the article systems attested in the world's languages. To do so, the frequencies of the different types of article systems in the sample are shown in Figure 9.1. Perhaps unsurprisingly, the most frequent systems consist of single articles. Out of 104 languages in the sample, 26 have a single definite article, 21 a single indefinite article, and 15 languages have a single anaphoric article. This distribution reflects the frequency distribution of article types in the sample shown in Section 8.1. Taken together, these $26+21+15=62$ languages make up almost 60% of all the languages in the sample. There may be a simple explanation for the fact that such simpler systems are more frequent than systems consisting of two or more articles. Each article is the result of a grammaticalization process. Therefore, systems that have more than one article have developed through more than one process of grammaticalization. Languages can generally do well without articles, i.e. the development of articles is in no way inherently necessary. Given that, the probability of finding more than one article in a single language can be expected to be lower than the probability of finding only one article, because two or more (independent) grammaticalization processes need to have taken place to result in a more complex article system. The observation that most languages in the sample have only a definite, anaphoric, or an indefinite article is strong evidence for the development of articles in the definite domain being largely independent from the development of articles in the indefinite domain, and vice versa.

The next frequent type of article systems is a system consisting of a definite and an indefinite article. It could be that this is a consequence of definite and indefinite articles being the two most frequent types of articles in the world's languages. However, given that anaphoric articles are comparatively common as well, we would expect to find that systems with an anaphoric and an indefinite article are more or less common as well. Such a system, however, only occurs in 2 languages of the sample, which calls for a different explanation. Furthermore, the preference for a DEF-INDEF system may not be a global preference. Figure 9.2 shows the occurrences of article systems in the sample across the six macro areas.

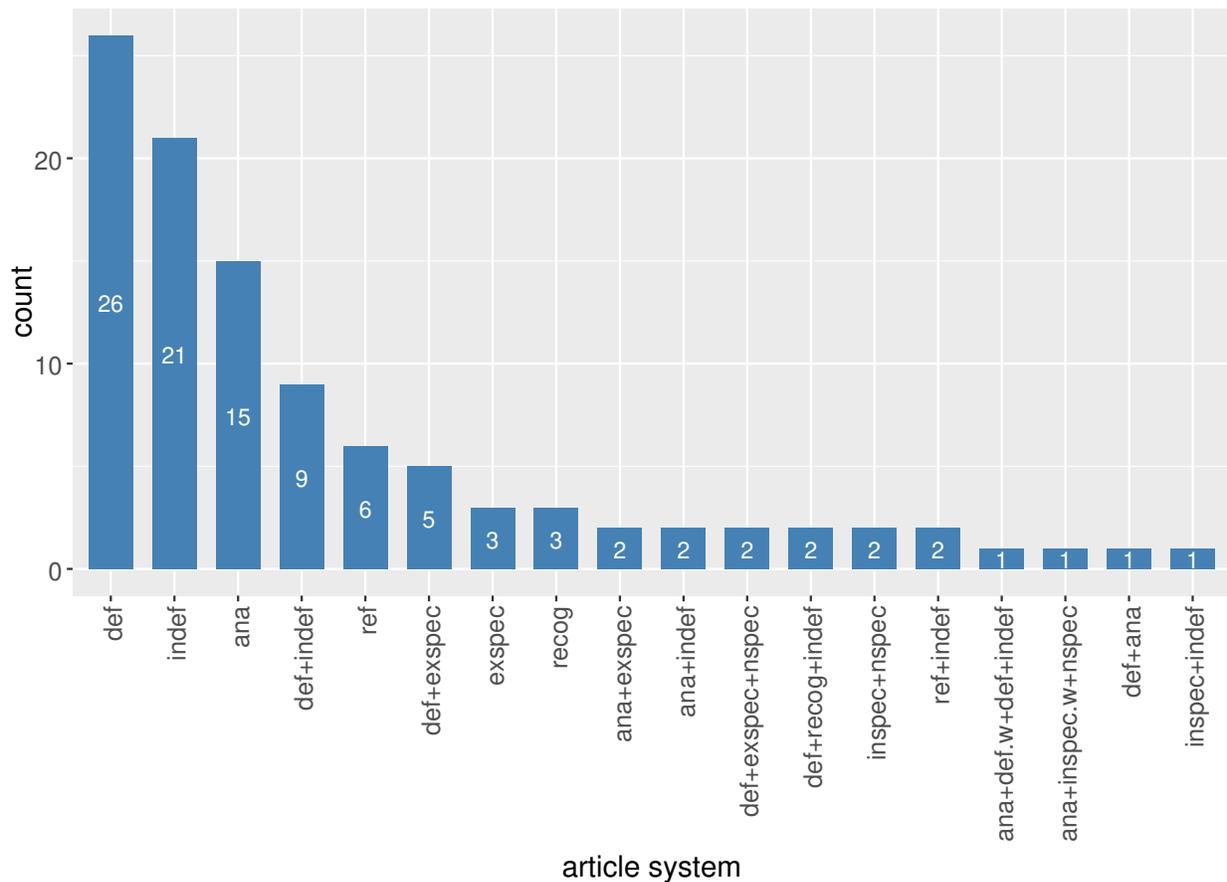


Figure 9.1: Distribution of article systems in the sample

The distribution partially reflects the areal trends seen for single articles (cf. Section 8.2) in that DEF-INDEF-systems occur in Eurasia and North America, but less so in the other macro areas. Especially Eurasia, besides South America, has a comparatively high proportion of indefinite articles. Definite articles being the most common type of articles in general, it may not be surprising that this leads to DEF-INDEF-systems. Interestingly, South America also has a rather high number of indefinite articles, but they occur mostly without any other articles in the system. It is thus very likely that areal effects due to contact and diffusion may have led to various areal biases towards

certain article systems. To what extent such extra-linguistic factors play a role in the crosslinguistic distribution of article systems goes beyond the scope of this study; further work will be needed to disentangle those different factors.

Also compared to other systems that cover the entire referential scale, the DEF-INDEF-system is more frequent than e.g. the INSPEC-NSPEC-system or the DEF-EXSPEC-NSPEC, each of which occur in 2 languages in the sample. However, the sample also contains 6 languages with a single referential article. Compared to 9 languages with a DEF-INDEF-system, REF-systems do not seem to be substantially less frequent across languages than the former.

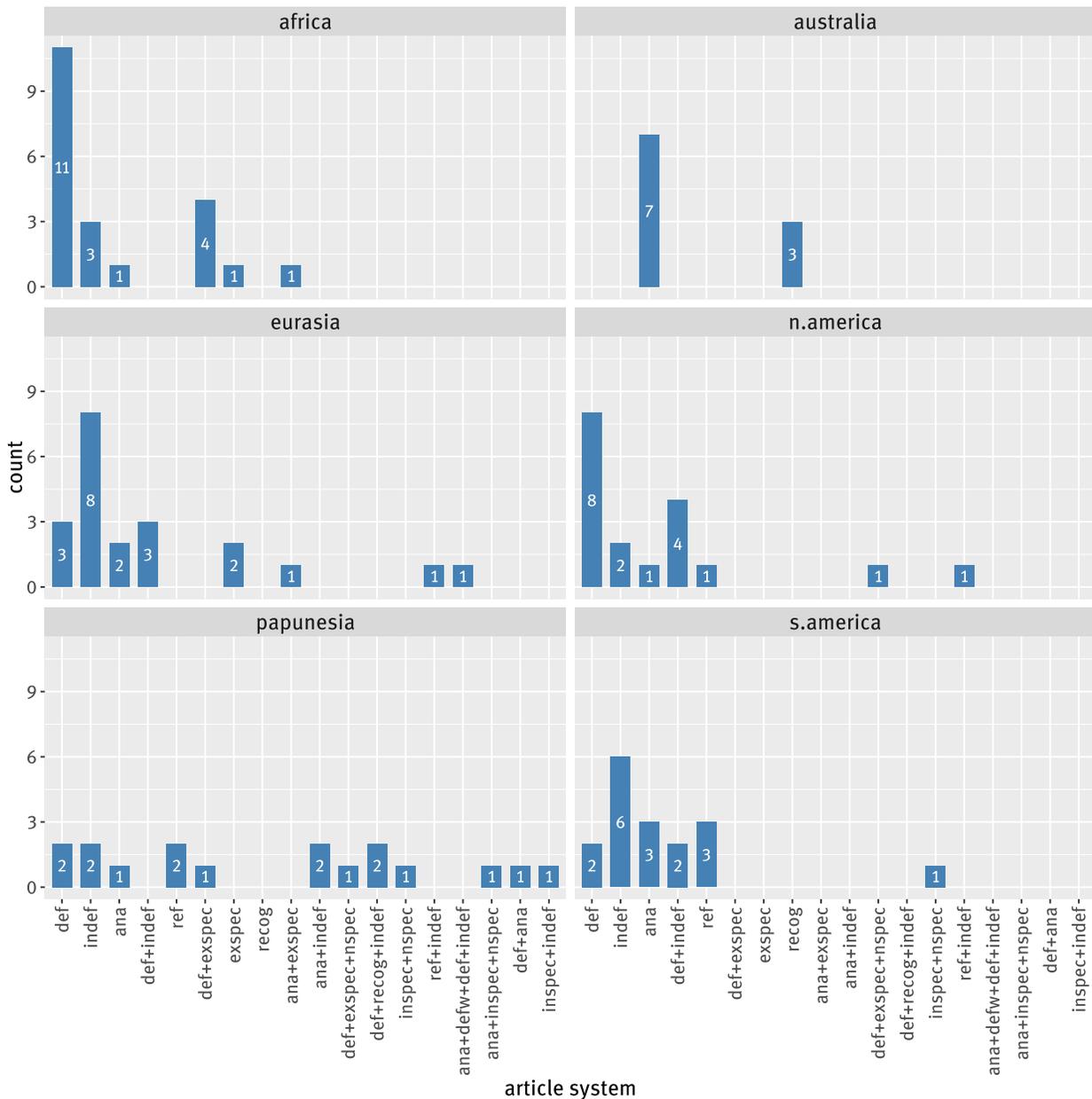


Figure 9.2: Distribution of article systems across macro areas

The last system with five or more occurrences in the sample is the system consisting of a definite and an exclusive-specific article, four of which are languages spoken in Africa. This points towards an areal bias as well. Other areal biases that can be seen in Figure 9.2 largely reflect the areal trends of single articles discussed in Section 8.2. Australia only has single anaphoric or recognitional articles, Africa and North America have the highest proportions of systems consisting of a single definite article, and South America as well as Eurasia have the highest proportions of systems with a single indefinite article. Interestingly, Papunesia has the greatest variety of different types of article systems in comparison with the other macro areas.

9.4.6 Evidence for a definite and an indefinite domain

The distributions of article systems presented in this chapter point towards a crosslinguistic preference to split the referential space, i.e. the various functions on the referential scale. This relates to the findings of Ionin (2006) based on a comparison of English and Samoan. While the theoretical assumptions of her approach are different from the ones in the present study, Ionin already noted that there are two possible ways in which the referential space can be split if the system has two articles. Although Ionin (2006: 233) uses slightly different labels, her distinction corresponds to INSPEC-NSPEC-systems vs. DEF-INDEF-systems (types 19 and 20) as defined in this study.

Ionin uses the existence of INSPEC-NSPEC-systems to argue that article systems can either make a split between the definite and the indefinite domain, or, they can group together definite and specific contexts as opposed to nonspecific contexts. This chapter showed that there are many other types of article systems that consist of two articles than these two. However, the opposition of these two systems results in an interesting observation, namely that the DEF-INDEF split is crosslinguistically more frequent than the INSPEC-NSPEC split.

The crosslinguistic preference for a split along the definite and the indefinite domain is not only found with those two article systems but with systems of different degrees of complexity. The relevant systems are shown in Table 9.7. For the purposes of this section, I only make a three-way distinction of referential contexts into definite, specific, and nonspecific. We can then compare how article systems consisting of 1, 2, or 3 articles, i.e. systems that are equally complex, split up the referential space. Here, I only show those systems that make either a DEF-INDEF or a INSPEC-NSPEC split. Of course, there are many other systems that do both, which are not shown here.

Table 9.7: DEF-INDEF and INSPEC-NSPEC splits in article systems

N articles	definite-indefinite split			in.specific-nonspecific split				
	R _{def}	R _{spec}	R _{nspec}	N	R _{def}	R _{spec}	R _{nspec}	N
1 article	DEF		no ART	26	no ART		NSPEC	0
	no ART		INDEF	21	INSPEC		no ART	(1)
2 articles	DEF		INDEF	9	INSPEC		NSPEC	2
3 articles	ANA+DEF _{WEAK}		INDEF	(1)	ANA+INSPEC _{WEAK}		NSPEC	1

Starting with systems that use one article, we can distinguish between a system with a single definite article as opposed to a system with a single nonspecific article (first row of Table 9.7). In the case of the former, specific and nonspecific referent types are left unmarked; in the case of the latter, definite and specific referents are not marked by any article. With 26 occurrences, the system with a single definite article corresponds to the most frequent article system in the sample. Systems of a single nonspecific article, on the other hand, are not attested crosslinguistically. The second row in Table 9.7 shows the two mirror systems, consisting of an indefinite article and an inclusive-specific article. Systems of a single indefinite article were shown to be the second most frequent type of article systems in the sample, being found in 26 out of 104 languages. Again, its counterpart system with a split between inclusive-specific and nonspecific is much less frequent. Such a system is found in Bemba, but not attested in any language of the sample. The third row shows systems with two articles, namely the traditional DEF-INDEF-system as opposed to the INSPEC-NSPEC-system. Again, the former system is more frequent than the latter. The last row in Table 9.7 shows the systems consisting of three articles that make a DEF-INDEF-split or an INSPEC-NSPEC-split. Both systems include a weak article which are so rare crosslinguistically that nothing can be concluded from the figures.

Summing up, the distribution of article systems in the sample shows that there is a crosslinguistic preference for the split of the referential space into a definite and an indefinite domain. This allows us to formulate Universal 7:

Universal 7 *The split into a definite and an indefinite domain*

In the world's languages, articles show a preference for a split of the referential space into a definite and an indefinite domain over a split into an inclusive-specific and a nonspecific domain.

One of the obvious reasons for this preference is that definite and indefinite articles are more frequent than inclusive-specific and nonspecific articles. The development of both of the latter article types can be expected to involve a more complex process than the development of definite or indefinite articles. Therefore, there may be no good synchronic or cognitive reason for the

preference of a DEF-INDEF-split, which may simply reflect the diachrony of the development of different article types.

9.5 Summary

This chapter gave a crosslinguistic overview of the article systems attested in the world's languages. Section 9.1 discussed the two cases of dependency between articles within single systems. One type of dependency was found with weak definite articles, which do not occur without anaphoric articles within the same system. Another type was found with nonspecific articles that are only attested in systems with either an inclusive-specific or an exclusive-specific article. In both cases, this dependency was argued to go back to the development of weak definite and nonspecific articles.

Sections 9.2 and 9.3 dealt with systems in which some articles were shown to have overlapping referential functions. Section 9.2 discussed systems with definite and anaphoric or recognitional articles. I argued that the functional overlap between a definite article and an anaphoric or recognitional one is similar to the functional overlap between a definite article and a demonstrative in any language with a definite article. Usually, the anaphoric article is used in contexts of topic shifts, when the antecedent is more distant and the referent more difficult to identify, and when it is important for the speaker to highlight that the referent is pragmatically definite.

The question of functionally overlapping articles in the indefinite domain was addressed in Section 9.3, presenting the article systems of Tepehua, Basque, and Maori. Tepehua and Basque were shown to have systems consisting of a referential article and an additional presentational article. In both languages, the presentational article was used with both specific and nonspecific referents in those cases in which the referential article was not sufficient to either mark a discourse-prominent referent as non-identifiable, or to emphasize the non-identifiability of the referent. The Maori system was shown to be a complex system with an inclusive-specific article, a presentational article, and a predicate marker that could also be used to mark indefinite referents. Besides a number of factors conditioning the use of the predicate marker, we saw that similar discourse-pragmatic factors to the ones in Tepehua and Basque influenced the use of the presentational article in Maori. It was shown to be used especially with discourse-prominent referents. Less prominent indefinite referents are encoded by the predicate marker. Interestingly, the data suggested that the inclusive-specific article is also used for specific referents of low discourse-prominence in those cases in which the use of the predicate marker was blocked for syntactic reasons.

Section 9.4 provided a detailed overview of attested and unattested article systems. On the basis of the 104-languages sample and a few additional languages discussed in the previous chapters,

I showed that we can distinguish between 22 types of article systems. The systems were grouped by how they cover the different referential functions on the referentiality scale: systems that only have articles in the definite domain, in the indefinite domain, in both domains, and systems with functionally overlapping articles. In addition, I discussed 9 types of article systems that are not attested. I argued for some of those that they cannot be categorically excluded but have to be assumed to be extremely infrequent across languages, while others could be excluded on the basis of the dependencies involving weak and nonspecific articles.

Section 9.4.5 briefly described the crosslinguistic distribution of those article systems. The distributions mostly mirrored the ones of single articles in that systems with single definite, indefinite, and anaphoric articles were the three most frequent types of article systems in the sample. With 26, 21, and 15 languages, respectively, they were shown to make up 60% of the systems in the sample of 104 languages. Article systems with two articles were shown to be comparatively rare across languages, the systems of DEF-INDEF and DEF-EXSPEC being the most frequent 2-article systems. The areal trends concerning article systems partially confirmed earlier results of articles as well. The DEF-INDEF-system was shown to be more common in Eurasia and North America than in the other macro areas. Africa proved to have the highest proportion of systems with a single definite article, also having comparatively many DEF-EXSPEC-systems. Eurasia and South America had the highest proportions of systems with indefinite articles, and Australia the highest number of anaphoric articles. Papunesia emerged as the area with the highest degree of variation in terms of article systems, with many of the less frequent systems in the sample being found in that area.

Section 9.4.6 pointed towards a general preference to split the referential space into a definite and an indefinite domain. Comparing the occurrence of systems with this type of split to systems with a split into inclusive-specific and nonspecific showed that systems of the latter type are much less frequent in the world's languages. I argued that this is most likely the result of the diachronic complexity of inclusive-specific and nonspecific articles rather than any synchronic preference.

Chapter 10

Conclusions

This study proposed a principled way to define and distinguish different types of articles on the basis of their referential function, presenting each of those article types in detail. Making use of a large sample, the study also offered new insights into the crosslinguistic distribution of article types and article systems.

10.1 Main findings

10.1.1 Article types

I showed in Chapter 2 that the first important step towards a crosslinguistic comparison of articles is a language-independent definition of articles. Defining articles as adnominal referential markers, Chapter 3 proposed a detailed solution. I first introduced the relevant referential functions and used those as the basis for the distinction of 10 different article types attested in the world's languages. The article types are: definite (DEF), anaphoric (ANA), weak definite (DEF_{WEAK}), recognitional (RECOG), indefinite (INDEF), exclusive-specific (EXSPEC), nonspecific (NSPEC), inclusive-specific (INSPEC), weak inclusive-specific (INSPEC_{WEAK}), and referential (REF) articles. Another type of articles are presentational articles (cf. Section 10.1.2), which are not listed here because they belong to indefinite articles in that they have the same referential functions. Table 10.1 shows these 10 article types, mapped to the referential functions that they encode.

Chapter 8 was concerned with the crosslinguistic distribution of articles. The three most frequent types of articles in the sample are definite (33% of the sample), indefinite (27%), and anaphoric (16%) articles. These figures show that definite articles should not be viewed as substantially more frequent than indefinite articles, even though they are the most frequent type. All other types of articles are less common and each only make up less than 10% of the sample. As for their distribution across the six macro areas of Africa, Australia, Eurasia, North Amer-

ica, Papunesia, and South America, five trends emerged from the data in the sample. Africa and North America appear to have a rather high proportion of definite articles, while Eurasia and South America showed the highest proportion of indefinite articles. Exclusive-specific articles appeared to be more common in Africa than in the other macro areas, and Australia and Papunesia were the only areas that recognitional articles were found in. Australia was discussed as a special case for being the macro area with the strongest overall trend against articles. I argued that a number of factors could have led to Australia having almost no articles except for a few anaphoric and recognitional ones. Complex demonstrative paradigms and the lack of distinct third person pronouns are likely to be a setting favoring the development of anaphoric articles from frequently used anaphoric pronouns. I argued that what has been described as determiners and potential articles in Australian languages often did not qualify as articles because their main functions were the structuring of the discourse rather than marking the referential function of the noun. In addition, flexible noun phrase structures that are found in a number of Australian languages were argued to be a syntactic property that may disfavor the development of articles. The present study also confirmed previous observations about the lack of generic articles, i.e. articles whose main function it would be to express kind reference. This was captured by Universal 1.

Universal 1 *The absence of generic articles*

There are no articles whose main referential function is the coding of generic referents.

Table 10.1: The 10 article types with their referential functions

article	definite domain							indefinite domain	
	R_{anaph}	R_{recog}	R_{establ}	$R_{rel-bridge}$	$R_{sit.unique}$	$R_{cont.unique}$	$R_{u-bridge}$	R_{spec}	R_{nspec}
DEF	✓	✓	✓	✓	✓	✓	✓		
ANA	✓	(✓)	(✓)	(✓)	(✓)				
DEF _{WEAK}					✓	✓	✓		
RECOG		✓	(✓)						
INDEF								✓	✓
EXSPEC								✓	
NSPEC									✓
INSPEC	✓	✓	✓	✓	✓	✓	✓	✓	
INSPEC _{WEAK}					✓	✓	✓	✓	
REF	✓	✓	✓	✓	✓	✓	✓	✓	✓

A by-product of classifying articles into types according to their referential functions and of discussing a variety of different examples of articles especially in the definite domain in Chapter 5, is

a refined referential scale. Building on previous work on referentiality, especially different areas of definiteness (Hawkins 1978, Himmelmann 1997, Löbner 1985, Schwarz 2009), the crosslinguistic discussion of articles could combine previous insights into the referential scale shown in (795). Especially the patterns of definite, anaphoric, weak definite, and recognitional articles showed that a much more fine-grained distinction of the definite domain into (at least) the referential functions proposed in this study is necessary in order to understand the use and distribution of these different types of articles. While the distinction into pragmatic definite and semantic definite contexts is not a new one, I showed that situational uniqueness must be distinguished from deictic contexts. Allowing for the construction of the referent as identifiable based on both its situational properties and based on its uniqueness in this situation, situational uniqueness can be constructed both as pragmatic definite and semantic definite. This was shown through the behavior of anaphoric and weak definite articles, which could both (at least in certain cases) mark referents as situationally unique.

(795) **Referential scale**

deictic > anaphoric > recognitional, establishing, rel-bridging >
situationally unique > contextually unique, u-bridging >
specific > nonspecific > non-referential

Another important result is the confirmation of earlier work that two different types of bridging contexts need to be distinguished; rel-bridging (relational bridging) as a subtype of anaphoric relations and u-bridging (unique bridging) as a subtype of contextual uniqueness. The data of different articles in the definite domain in this study confirmed that u-bridging indeed appeared to be marked by articles in the same way as referents that are contextually unique, i.e. by definite and weak definite articles. However, rel-bridging was not necessarily encoded by articles in the same way as anaphoric contexts were. Hence, more typological and theoretical research is needed to shed more light on the properties of especially rel-bridging contexts.

Related to the referential scale, the distribution of article types analysed in Chapter 8 showed that most articles encode pragmatic definite referents, while semantic definite, specific, and non-specific referents are encoded by fewer articles across languages. Although I showed that definite articles are not necessarily the most frequent type of articles across languages, the probability of articles encoding pragmatic definite expressions is indeed higher than the probability of encoding other types of referential functions. This insight was captured by Universal 2.

Universal 2 *The expression of referential functions by articles*

The probability of referential functions being encoded is highest for pragmatic definites and lower for semantic definite, specific, and nonspecific functions.

Related to that, the types of articles and systems attested suggested that articles generally extend their referential functions along the referential scale from pragmatic definite, semantic definite, specific indefinite, and finally nonspecific indefinite. This gradual diachronic extension is plausible because we find anaphoric, definite, inclusive-specific, and referential articles. At the same time, no evidence for a development in the opposite direction was found. I described one notable exception, namely possessive markers that extend their functions to pragmatic definite contexts. I showed that these markers are not articles, and that they do not primarily undergo an extension of their referential functions but of their discourse-pragmatic functions. This directed development of articles was captured in Universal 3. The article types that appear to be counter-evidence, e.g. weak definite or nonspecific articles, were shown to result from an additional later restriction of a former definite or indefinite article.

Universal 3 *The functional extension of articles*

Articles generally extend their functions along the referential scale, from pragmatic definite to semantic definite to specific and to nonspecific contexts.

Another important insight relating to the referential functions of articles and motivating the structure of the referential scale was the observation that articles only encode adjacent functions on the scale. This led to the formulation of Universal 4.

Universal 4 *Articles encode adjacent functions on the referential scale*

Articles do not encode functions on the referentiality scale that are not adjacent. This excludes the existence of, for instance, an article that marks definite and nonspecific referents without marking specific referents.

10.1.2 Articles and discourse prominence

In Chapter 6, I showed that we find a group of indefinite articles in the world's languages whose use is restricted to discourse-prominent referents. I called this type of indefinite article presentational article. The existence of presentational articles has an important consequence for the widely assumed grammaticalization path from the numeral 'one' to an indefinite article as proposed in the literature. It has been assumed that the numeral first develops into a presentative marker, introducing new discourse referents, and then into a specific article, occurring with specific referents regardless of their discourse prominence. Only then the specific marker extends its use to nonspecific contexts, reaching the stage of an indefinite article, so the argument goes. The presentational articles discussed in Section 6.2, however, showed that the relation between the semantic extension from specific to nonspecific contexts and the pragmatic extension from discourse-prominent to less prominent referents are independent from each other. Like

indefinite articles, presentational articles were shown to encode specific and nonspecific referents. However, they are subject to further pragmatic restrictions in that they are only used with discourse-prominent referents. This is counter-evidence to a strict joined semantic and pragmatic extension because the use of presentational articles are not restricted to specific indefinite contexts. This means that articles can extend their referential functions towards nonspecific contexts without necessarily losing their pragmatic restriction. Moreover, the sample contained about as many presentational articles (18) as indefinite articles (19), which strongly suggested that presentational articles are not merely an intermediate stage towards indefinite articles, and that they should be viewed as a type of article in their own right.

Presentational articles were also found to play an important role in article systems with a functionally broad inclusive-specific or referential article. In Chapter 9, I presented the article systems of Basque, Huehuetla Tepehua, and Maori, showing how an inclusive-specific or referential article is replaced by a presentational article in those contexts in which a discourse-prominent referent requires to be signalled as non-identifiable. Because a discourse-prominent referent is interpreted as definite by default, and because an inclusive-specific or referential article cannot disambiguate between a definite or a specific interpretation of the referent, an additional presentational article is used instead in some languages. This showed that the use and development of articles is clearly also driven by pragmatic factors.

Another related observation from the definite domain in Chapter 5 was the availability of anaphoric markers in addition to articles whose use is also pragmatically restricted. As a minor side note, I showed examples from Hausa and Lakota, where those anaphoric markers are used when the distance to the antecedent is large, often also to signal that the current referent is a re-activated topic. I argued that such markers should not be classified as articles, since their use is very similar to demonstratives in general. Thus, while the nature of the discourse-pragmatic restrictions in the case of “topical” anaphoric markers is slightly different from those of presentational articles, we equally see discourse-pragmatic effects or constraints in the definite and the indefinite domain. Examining those factors in detail, and including referential markers that are pragmatically restricted in this way regardless of being or articles or not was beyond the scope of this study. This is certainly an issue that requires more research. I also used the use of the notion of discourse prominence in a very simplistic way without going into more detail with respect to the exact contextual and inherent conditions that contribute to a referent being more or less discourse-prominent, and how that restricts the use of the article in single languages. The inherent semantic properties of the referent such as animacy as well as its status as the center of attention in the immediately following discourse segment are most likely two important factors. A proper account however would require a more detailed corpus-based investigation of such markers within single languages. Even though preliminary, the results of this study do point to the fact that such

restrictions of articles and other determiners are rather common across the world's languages. Also the situation in Australia, briefly mentioned in Chapter 8, where referential markers are not common but where various linguistic devices are commonly used to express discourse-pragmatic functions, calls for a closer look at how the domains of articles and discourse-structuring devices interact and complement each other.

10.1.3 Article systems

Chapter 9 dealt with article systems from a crosslinguistic perspective. To the best of my knowledge, this has not yet been done in such a systematic way using a large sample. Based on the languages in the sample and a few additional examples discussed in Chapters 5, 6, and 7, I showed that we need to distinguish 22 types of attested article systems. The most frequent systems in the sample consisted of a single definite, indefinite, or anaphoric article. With 26, 21, and 15 out of 104 languages in the sample each, these three systems comprise almost 60% of the sample. From the systems with two articles, the most frequent type was that of a definite article together with an indefinite article, occurring in 9 languages of the sample. Article systems consisting of more than two articles were shown to be crosslinguistically rare. Very complex attested systems were a system with a definite, an exclusive-specific, and a nonspecific article, as well as a system with an anaphoric, a weak inclusive-specific, and a nonspecific article. The only other languages in the sample with such complex systems have basic system of a definite and an indefinite article together with an additional recognitional or weak definite article.

In general, the distributions of articles across different types of systems suggested that there is not a high degree of dependency between articles in the definite domain and articles in the indefinite domain. In other words, the study found systems with articles in only one of the two domains as well as in both domains. However, I showed that we do find two important types of dependencies between articles within the same system. This dependency involved weak definite (as well as the other weak articles) and anaphoric articles in the definite domain, and nonspecific and exclusive/inclusive-specific articles in the indefinite domain. I argued that both types of dependency can be explained by the diachronic processes that lead to their development. Weak definite articles only occur in systems that also include an anaphoric article because weak definite articles only arise in systems where the anaphoric article gradually takes over the functions of the definite article in pragmatic definite contexts. In German, for instance, this may not necessarily have been a proper anaphoric marker as such, but the system had the option of a phonetically longer and a phonetically reduced marker in certain contexts, in which the longer marker, being more salient, became conventionalized and developed into an anaphoric article. If the latter is then so strongly associated with pragmatic definite contexts, it may block the use of the former definite article in those contexts, making the latter a weak definite article. I showed that the same

holds for what I tentatively analysed as a weak inclusive-specific article in Tongan. While weak referential articles are not attested to the best of my knowledge, they may also exist, in which case their dependency on an anaphoric article was predicted to be the same as for the other two weak article types. This dependency was captured by Universal 5.

Universal 5 *Weak articles depend on anaphoric articles*

Weak definite, weak inclusive-specific, and weak referential articles only occur in article systems that include an anaphoric article as well.

Such weak articles are extremely rare crosslinguistically; I showed in Chapter 5 that most of the presumed weak definite articles mentioned in the literature are in fact not blocked in pragmatic definite contexts in general. Rather, pragmatic factors seem to condition whether the definite article is sufficient in such contexts or whether a “stronger” anaphoric article is used.

I showed that we find a similar dependency between nonspecific and exclusive/inclusive-articles. I argued that this dependency can equally be explained by the diachronic processes through which a nonspecific article arises. Although more work is needed to confirm the details of the processes presented in this study, nonspecific articles were shown to originate from indefinite articles. I argued that indefinite articles can be restricted in their use to nonspecific contexts, for instance in systems in which a former definite article extends its function towards marking specific indefinite referents. Why and how such a development takes place will require future work; but the extension of definite articles towards the indefinite domain is attested in a number of languages. In that case, the former indefinite article may become a nonspecific article. This scenario was argued to have happened in a number of Oceanic languages with nonspecific articles. Another scenario that can lead to the development of nonspecific articles (or markers) was argued to be the extended use of irrealis markers with nonspecific referents. I showed that in a number of Central and North American languages, we find that irrealis markers either replace the indefinite article in nonspecific contexts or that they are used in addition, forming a new nonspecific marker. If the former indefinite article is no longer felicitous in nonspecific contexts, it thus developed into an exclusive-specific article. Because the nonspecific article in those systems originates from an indefinite article whose use in specific contexts is restricted, I argued that nonspecific articles do not occur in article systems without exclusive/inclusive-specific articles. This was expressed as Universal 6.

Universal 6 *Nonspecific articles depend on specific articles*

Nonspecific articles only occur in article systems with either inclusive-specific or exclusive-specific articles.

Another trend regarding article systems observed relates to how the referential space is split up by the use of articles. The referential space was understood as all the functions on the referential

scale. I showed that we find a crosslinguistic preference for a split along the definite and the indefinite domain rather than between an inclusive-specific and a nonspecific domain. This trend was observed across article systems of varying degrees of complexity. For instance, the sample contained 26 languages with a single definite and 21 languages with a single indefinite article as opposed to no language with a single nonspecific article and one language with a single inclusive-specific article. However, I argued that this preference, formulated as Universal 7, is likely only a consequence of the diachrony of articles and article systems.

Universal 7 *The split into a definite and an indefinite domain*

In the world’s languages, articles show a preference for a split of the referential space into a definite and an indefinite domain over a split into an inclusive-specific and a nonspecific domain.

10.2 Future research

One area where more research is needed but for which this study has laid important groundwork is an even more fine-grained distinction of referential functions or combinations thereof. The present study has focused on the use of articles in rather simple referential contexts. However, there are of course more complex scenarios, e.g. non-identifiable referents that are mentioned again in a certain discourse situation, making them anaphoric at the same time. In Akan, for instance, we find an interesting interaction between the use of the anaphoric article *no* and the exclusive-specific article *bi*. As is shown in example (796), both articles can be used together to mark a non-identifiable specific referent that was previously mentioned.

- (796) [[maame **bi**] **no**] a ɔ-bɛ-hwehwɛ-ɛ wo no ...
 woman ART:EXSPEC ART:ANA REL 3SG-MP-look.RED-COMPL you CD ...
 ‘That (certain) woman who came looking for you ...’
 Akan (Amfo 2010: 1797)

Interestingly, the two articles in Akan can also be combined in the reverse order and scope relation; example (797) shows the noun *nkɔrɔfo* ‘people’ with both an anaphoric and an exclusive-specific article, and we can see in the translation that this combination has a partitive interpretation.

- (797) [[nkɔrɔfo **no**] **bi**] ka-a sɛ wɔ.re-m-pene
 people ART:ANA ART:EXSPEC say-COMPL COMP 3PL-PROG-NEG-agree
 ‘Some of the people said that they will not agree.’
 Akan (Amfo 2010: 1795)

In general, although not very common, the co-occurrence of articles is attested in other languages as well, often expressing an additional function or meaning. Also in Tongan, for instance, MacDonald (2014: 88-89) provides examples of the nonspecific article co-occurring with the so-called definite accent, which I analysed as part of the anaphoric article. One such example is shown in (798). However, MacDonald (2014: 88) also notes that the combination of the nonspecific article *ha* and the definite accent is only felicitous in the presence of the emphatic particle *pē*.

- (798) Te u fiemālie pē au ki [ha taimi pē ‘e faingamālie kiate
 SUBJ 1EX:SG content EMPH 1EX:SG DAT ART:NSPEC time EMPH SUBJ convenient DAT
 koé].
 2SG.ART:ANA
 ‘I will be content only with a time that is convenient to you.’
 Tongan (MacDonald 2014: 89)

Thus, future research is needed to draw a clearer picture of how common such co-occurring articles are across the world’s languages, and what their referential (or other) functions are in such constructions.

Returning to the point about more complex referential scenarios, another important type of contexts needs to be mentioned, namely so-called non-referential definite expressions (cf. Donnellan 1966, von Heusinger 2002) shown in example (799). In well-known languages like English with a DEF-INDEF-system, the definite article is used in both types of expressions in (799). In languages with an INSPEC-NSPEC-system, however, one may expect that those two interpretations are overtly marked, using the inclusive-specific article in contexts like (799a) and the nonspecific article in contexts like (799b) (cf. Ionin 2006: 218). In fact, the Tongan example shown above in (798) could be such a context in which the referent is marked as nonspecific for the speaker and as definite for the hearer at the same time.

- (799) John is looking for the dean.
 a. ..., whoever it might be. (non-referential)
 b. ..., namely for Smith, who happens to be the dean. (referential)
 (von Heusinger 2002: 248)

Another aspect of articles that has emerged from the patterns discussed in this study is that articles are often sensitive to discourse-pragmatic factors. The results of this study also suggested that the factors are fairly similar across languages, but how exactly the distribution of articles is driven by such factors will require quantitative or comparative studies of different article systems in the future.

As for article systems, I presented certain cases in which the presence of one article depended on the presence on another article in the system. In most cases, however, the results of this study

suggested that articles from the definite and the indefinite domain emerge fairly independently from each other. However, the question of how the presence or absence of one type of articles can influence the emergence, stability, or loss of another article in the same language was only touched upon and deserves a closer look in future research. The same holds for other grammatical properties that are very likely to have an impact on the presence or absence of articles. The most obvious candidates are other grammatical devices that serve to introduce and to track discourse referents, e.g. verbal agreement, topic and focus markers, obviation systems, noun class systems with class-marked pronouns, etc. We know that the loss of case marking had an impact on the development of article systems in European languages. Impressionistically, there does not seem to be any general crosslinguistic relation between article systems and case marking, and the availability of articles and their interactions with other grammatical categories appear to be subject to areal biases. Further research is needed to test to what extent this is indeed the case. A related matter is the question of areal biases towards or against the presence of certain types of articles. I presented a few areal trends that emerged from the distribution in the sample, but shedding more light on this question will require a different approach. First, one would need to include languages without articles, which were categorically excluded in the present study. Second, given the complexity of factors that articles interact with, one may need to examine specific areas in order to be able to capture contact and diffusion and language-internal processes over time in sufficient detail. In a similar vein, comparing languages with and without articles will allow to examine the tools that languages without articles use to express the functions of articles presented in this study.

To conclude, there are still a number of unanswered questions, especially regarding additional factors that condition the presence or absence of articles, other grammatical properties that articles interact with, and the functions that articles have in addition to their referential functions. Analysing the data of a large-scale crosslinguistic sample, this study established a sound overview of articles and their referential functions, proposing a typology of articles and article systems. Hopefully, this contribution to our understanding of articles will serve as a point of departure for many studies of articles in the future.

Language Index

- Agta (Dupaningan), 365
Aguaruna, 172
Akan, 160–170, 242–243
Albanian, 42
Amharic, 114–115
Angolar, 362
Apinayé, 146
Arawak, 147, 150
Armenian, 38, 309
Arrernte, 173
Ayoreo, 122, 125, 269–271
Ayutla Mixe, 117

Bambassi, 147
Basque, 310, 341–347
Baure, 283–287, 310
Bemba, 35, 265–269, 310
Biak, 148, 151
Bilua, 227–229
Bininj Kun-Wok, 202–204
Bonan, 221–224, 311
Bullom So, 38, 142, 309
Bunaq, 171
Buwal, 141

Ch’ol, 147, 260
Chatino (Teotepéc), 229–231
Chinese, 29
Coptic, 47

Crow, 46, 261

Dime, 142, 146

Ejagham, 106
English, 35, 45, 334

Fering, 174–181

German, 29, 44, 151, 171, 174–181
Gooniyandi, 207–209

Halkomelem, 287–293
Hausa, 43, 181–186
Hidatsa, 260, 332

Icelandic, 331
Indonesian, 115–117
Irish, 150

Jamsay, 328

Kaqchikel, 129–133, 142–145, 330
Kashibo-Kakataibo, 231–235
Khwarshi, 40
Koasati, 37
Komnzo, 157–160
Konso, 46

Lakota, 186–190, 248–250, 262
Lango, 225–227
Lavukaleve, 146, 199–202, 334

Limbum, 106, 152–157
Ma'di, 193–195
Makah, 262
Maori, 347–360
Mocho, 260
Mokpe, 39, 45, 134–141, 146
Mongsen Ao, 173

Nivaclé, 30
Nuuchahnulth, 141, 262
Nyaturu, 39

Oko, 310
Oksapmin, 107, 196–199, 334

Palula, 121, 243, 244, 311

Q'anjob'al, 47, 257–260, 332

Rapa Nui, 151, 278–283

Sheko, 45, 151, 310

Siar Lak, 332
Slovenian, 40
Sri Lanka Malay, 39, 218–221
Sumu, 38

Tarahumara, 150
Teiwa, 171, 364
Tepehua, 38, 336–340
Tikuna, 120
Tongan, 251–257, 272–278
Turkish, 38
Tz'utujil, 48, 216–218, 324

Udmurt, 29, 111–114
Ughele, 36
Ulwa, 142
Urama, 190–193, 334
Ute, 38

Yankunytjatjara, 204–207

Bibliography

- Abbott, Barbara. 2010. *Reference*. Oxford: Oxford University Press.
- Abner, Natasha & Heather Burnett. 2010. *On the so-called “determiners” of Tongan*. UCLA. Unpublished Manuscript.
- Aboh, Enoch Oladé. 2005. Deriving relative and factive construction in Kwa. In Laura Brugè et al. (eds.), *Contributions to the thirtieth Incontro di Grammatica Generativa*, 265–285. Venezia: Libreria Editrice Cafoscarina.
- Aguilar-Guevara, Ana, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.). 2019. *Definiteness across languages* (Studies in Diversity Linguistics 25). Berlin: Language Science Press.
- Aguilar-Guevara, Ana & Joost Zwarts. 2010. Weak definites and reference to kinds. *Proceedings of SALT 20*. 179–196.
- Aguilar-Guevara, Ana & Joost Zwarts. 2013. Weak definites refer to kinds. *Recherches linguistiques de Vincennes* 42. 33–60.
- Ahland, Michael. 2012. *A grammar of Northern Mao (Màwés Aas’é)*. Eugene, OR: University of Oregon dissertation.
- Ahn, Byron. 2016. Syntax-phonology mapping and the Tongan DP. *Glossa* 1(1). 4.
- Aikhenvald, Alexandra Y. 1995. *Bare*. München: Lincom Europa.
- Aikhenvald, Alexandra Y. 2008. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. & Robert Malcolm Ward Dixon. 2013. *Possession and ownership*. Oxford: Oxford University Press.
- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21(3). 435–483.
- Álvarez, José. 2017. *Compendio de la gramática de la lengua Wayuu [Compendium of the grammar of the Wayuu language]*. Universidad del Zulia. Unpublished Manuscript.
- Am-David, Assif. 2016. *Semantic typology of languages with multiple definite articles*. Goethe-Universität Frankfurt am Main. Unpublished Manuscript.
- Ameka, Felix Kofi. 1991. *Ewe: Its grammatical constructions and illocutionary devices*. Canberra: Australian National University dissertation.

- Amfo, Nana Aba Appiah. 2010. Indefiniteness marking and Akan *bi*. *Journal of Pragmatics* 42(7). 1786–1798.
- Anderson, Stephen R. & Edward L. Keenan. 1985. Deixis. In *Language typology and syntactic description III: Grammatical categories and the lexicon*, 259–308. Cambridge: Cambridge University Press.
- Anderson, Victoria & Yuko Otsuka. 2006. The phonetics and phonology of “definitive accent” in Tongan. *Oceanic Linguistics* 45(1). 21–42.
- Antunes de Araujo, Gabriel. 2004. *A grammar of Sabanê. A Nambikwaran language*. Utrecht: LOT.
- Appleyard, David. 2005. Definite markers in modern Ethiopian Semitic languages. In Geoffrey Khan (ed.), *Semitic studies in honour of Edward Ullendorff*, 51–61. Leiden: Brill.
- Aragon, Carolina Coelho. 2014. *A grammar of Akuntsú, a Tupían language*. Manoa: University of Hawai’i Press.
- Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24(1). 65–87.
- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Ariel, Mira. 2001. Accessibility theory: An overview. In Ted J.M. Sanders, Joost Schilperoord & Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects*, 29–87. Amsterdam: Benjamins.
- Arkoh, Ruby. 2011. *Semantics of Akan bi and nu*. Vancouver: University of British Columbia dissertation.
- Arkoh, Ruby & Lisa Matthewson. 2013. A familiar definite article in Akan. *Lingua* 123. 1–30.
- Atindogbe, Gratien. 2013. *A grammatical sketch of Mòkpè (Bakweri) (Bantu A20)*. Kyoto: The Center for African Area Studies of Kyoto University.
- Atoyebi, Joseph Dele. 2010. *A reference grammar of Oko*. Köln: Köppe.
- Auer, Peter. 1981. Zur indexikalitätsmarkierenden Funktion der demonstrativen Artikelform in deutschen Konversationen. In Götz Hindelang & Werner Zillig (eds.), *Sprache: Verstehen und Handeln*, 301–311. Tübingen: Niemeyer.
- Auer, Peter. 1984. Referential problems in conversation. *Journal of Pragmatics* 8(5). 627–648.
- Baker, Brett. 2008. The interpretation of complex nominal expressions in Southeast Arnhem Land languages. In Ilana Mushin & Brett J. Baker (eds.), *Discourse and grammar in Australian languages*, 135–166. Amsterdam: Benjamins.
- Baker, Carl Lee. 1970. Double negatives. *Linguistic Inquiry* 1(2). 169–186.
- Baker, Carl Lee. 1973. *Definiteness and indefiniteness in English*. Bloomington, IN: Indiana University Linguistics Club.
- Baptista, Marlyse. 2002. *The syntax of Cape Verdean Creole: The Sotavento varieties*. Amsterdam: Benjamins.

- Barker, Chris. 2004. Possessive weak definites. In Ji-Yung Kim, Yury A. Lander & Barbara H. Partee (eds.), *Possessives and beyond: Semantics and syntax* (University of Massachusetts Occasional Papers in Linguistics 29), 89–113. Amherst, MA: GLSA.
- Barros Brandão, Ana Paula. 2014. *A reference grammar of Paresi-Haliti (Arawak)*. Austin, TX: University of Texas dissertation.
- Bauer, Winifred. 1993. *Maori*. London: Routledge.
- Beck, David. 2013. Unidirectional flexibility and the noun–verb distinction in Lushootseed. In Jan Rijkhoff & Eva Helena van Lier (eds.), *Flexible word classes typological studies of underspecified parts of speech*, 185–220. Oxford: Oxford University Press.
- Bertinetto, Pier Marco. 2009. Ayoreo (Zamuco). A grammatical sketch. *Quaderni del Laboratorio di Linguistica della Scuola Normale Superiore Pisa* 8. 1–59.
- Bhat, D.N.S. 2004. *Pronouns*. Oxford: Oxford University Press.
- Bhat, D.N.S. 2013. Third person pronouns and demonstratives. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Bisle-Müller, Hansjörg. 1991. *Artikelwörter im Deutschen*. Tübingen: Niemeyer.
- Blackings, Mairi & Nigel Fabb. 2003. *A grammar of Ma'di*. Berlin: De Gruyter.
- Blake, Barry J. 1983. Structure and word order in Kalkatungu: The anatomy of a flat language. *Australian Journal of Linguistics* 3. 143–175.
- de Blois, Kornelis Frans. 1970. The augment in the Bantu languages. *Africana Linguistica* 4. 85–165.
- Bohn, Ocke-Schwen. 2004. How to organize a fairly large vowel inventory: The vowels of Fering (North Frisian). *Journal of the International Phonetic Association* 34(2). 161–173.
- Bombi, Carla. 2018. Definiteness in Akan: Familiarity and uniqueness revisited. *Semantics and Linguistic Theory* 28(0). 141–160.
- Bombi, Carla et al. 2019. The semantics of the (so-called) clausal determiner *nó* in Akan (Kwa). In M. Theresa Espinal et al. (eds.), *Proceedings of Sinn und Bedeutung 23*, 181–199. Bellaterra: Universitat Autònoma de Barcelona.
- Boro, Krishna. 2017. *A grammar of Hakhun Tangsa*. Eugene, OR: University of Oregon dissertation.
- Bowern, Claire. 2004. *Bardi verb morphology in historical perspective*. Cambridge, MA: Harvard University dissertation.
- Bowern, Claire. 2008. Bardi arguments: Referentiality, agreement, and omission in Bardi discourse. In Ilana Mushin & Brett J. Baker (eds.), *Discourse and grammar in Australian languages*, 59–86. Amsterdam: Benjamins.
- Bowern, Claire. 2012. *A grammar of Bardi*. Berlin: De Gruyter.

- Brandstetter, Renward. 1913. *Der Artikel des Indonesischen verglichen mit dem des Indogermanischen*. Lucerne: E. Haag.
- Brown, Jason & James J. Thompson. 2013. The emergence of determiner clisis in Upriver Halkomelem. *Northwest Journal of Linguistics* 7(1). 1–13.
- Brown, Jason et al. 2016. *A short grammar of Urama*. Canberra: Asia-Pacific Linguistics.
- Brown, R. McKenna et al. 2006. *¿La üt̥z awäch? Introduction to Kaqchikel Maya language*. Austin, TX: University of Texas Press.
- Brugger, Gerhard & Martin Prinzhorn. 1996. *Some properties of German determiners*. Vienna. Unpublished Manuscript.
- Brugmann, Karl. 1904. *Die Demonstrativpronomina der indogermanischen Sprachen: Eine bedeutungsgeschichtliche Untersuchung*. Leipzig: Teubner.
- Buba, Malami. 1997. The deictic particle DI-N in Hausa. *African Languages and Cultures* 10(1). 29–45.
- Buchholz, Oda & Wilfried Fiedler. 1987. *Albanische Grammatik*. Leipzig: Verlag Enzyklopädie.
- Buechel, Eugen. 1939. *A grammar of Lakhota. The language of the Teon Sioux Indians*. Rosebud: Rosebud Educational Society.
- Bühler, Karl. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.
- Butterworth, George. 1998. What is special about pointing in babies? In Francesca Simion & George Butterworth (eds.), *The development of sensory, motor and cognitive capacities in early infancy. From perception to cognition*, 171–190. Hove: Psychology Press.
- Campbell, Lyle. 1985. *The Pipil language of El Salvador* (Mouton Grammar Library 1). Berlin: De Gruyter.
- Campbell, Lyle. 2012. Typological characteristics of South American indigenous languages. In Lyle Campbell & Verónica Grondona (eds.), *The Indigenous languages of South America. A comprehensive guide*, 259–330. Berlin: De Gruyter.
- Campbell, Sharon Stark & Verónica Luna Villanueva. 2011. *Gramática ngigua de San Marcos Tlacoyalco [Ngigua grammar of San Marcos Tlacoyalco]*. México, D.F.: Instituto Lingüístico de Verano, A.C.
- Carlson, Gregory & Francis J. Pelletier (eds.). 1995. *The generic book*. Chicago: University of Chicago Press.
- Carlson, Gregory N. 2005. Generics, habituals and iteratives. In Alex Barber (ed.), *Encyclopedia of language and linguistics*. Amsterdam: Elsevier.
- Carlson, Gregory N. & Rachel Shirley Sussman. 2005. Seemingly indefinite definites. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives* (Studies in Generative Grammar 85), 71–86. Berlin: De Gruyter.
- Carlson, Robert. 1994. *A grammar of Supyire* (Mouton Grammar Library 14). Berlin: De Gruyter.

- Carnap, Rudolf. 1947. *Meaning and necessity: A study in semantics and modal logic*. Chicago: University of Chicago Press.
- Chacon, Thiago Costa. 2012. *The phonology and morphology of Kubeo: The documentation, theory, and description of an Amazonian language*. Manoa: University of Hawai'i at Mānoa dissertation.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.
- Chen, Ping. 2003. Indefinite determiner introducing definite referent: A special use of 'yi 'one' + classifier' in Chinese. *Lingua* 113. 1169–1184.
- Chen, Rong. 1990. English demonstratives: A case of semantic expansion. *Language Sciences* 12(2). 139–153.
- Chiaros, Christian, Berry Claus & Michael Grabski (eds.). 2011. *Saliency* (Trends in Linguistics. Studies and Monographs 227). Berlin: De Gruyter.
- Chierchia, Gennaro. 1998a. Plurality of mass nouns and the notion of “semantic parameter”. In Susan Rothstein (ed.), *Events and grammar*, 53–103. Dordrecht: Kluwer.
- Chierchia, Gennaro. 1998b. Reference to kinds across languages. *Natural Language Semantics* 6(4). 339–405.
- Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention*. Oxford: Oxford University Press.
- Childs, George Tucker. 2011. *A grammar of Mani* (Mouton Grammar Library 54). Berlin: De Gruyter.
- Chonay Chonay, Ana María. 2006. *Runuk'ulem pa rub'eyal rutz'ib'axik ri kaqchikel ch'ab'al*. *Gramática normativa del idioma maya Kaqchikel*. [Normative grammar of the Mayan language Kaqchikel]. Guatemala: Academia de Lenguas Mayas de Guatemala.
- Christophersen, Paul. 1939. *The articles: A study of their theory and use in English*. Copenhagen: Einar Munksgaard.
- Chung, Sandra & William Allen Ladusaw. 2004. *Restriction and saturation*. Cambridge, MA: MIT Press.
- Chung, Sandra, Te Haumihiata Mason & J. W. Milroy. 1995. On Maori he and the uses of indefinites. *The Journal of the Polynesian Society* 104(4). 429–459.
- Churchward, C. Maxwell. 1953. *Tongan grammar*. Oxford: Oxford University Press.
- Churchward, C. Maxwell. 1985. *Tongan grammar*. Tonga: Vava'u Press.
- Ciucci, Luca. 2016. *Inflectional morphology in the Zamucoan languages*. Asunción: Biblioteca Paraguaya de Antropología.
- Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.

- Clark, Herbert H. & Susan E. Brennan. 1987. Grounding in communication. In Lauren B. Bresnick, John M. Levine & Stephanie D. Teasley (eds.), *Perspectives on socially shared cognition*, 127–149. Washington: American Psychological Association.
- Clark, Herbert H. & Susan E. Haviland. 1977. Comprehension and the given-new contract. In Roy O. Freedle (ed.), *Discourse production and comprehension*, 1–40. Norwood, NJ: Ablex.
- Clark, Ross. 1974. On the origin of the Tongan definitive accent. *The Journal of the Polynesian Society* 83. 103–108.
- Clendon, Mark. 2014. *Worrorra: A language of the north-west Kimberley coast*. Adelaide: University of Adelaide Press.
- Cohen, Diana. 1998. *A grammatical description of Tarahumara*. Arlington, TX: The University of Texas at Arlington dissertation.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*. 2nd edn. Oxford: Blackwell.
- Coupe, Alec R. 2007. *A grammar of Mongsén Ao* (Mouton Grammar Library 39). Berlin: De Gruyter.
- Courtz, Hendrik. 2008. *A Carib grammar and dictionary*. Toronto: Magoria Books.
- Cristofaro, Sonia. 2009. Grammatical categories and relations: Universality vs. language-specificity and construction-specificity. *Language and Linguistics Compass* 3(1). 441–479.
- Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.
- Croft, William. 2003. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press.
- Crowley, Terry. 1998. *An Erromangan (Sye) grammar*. Honolulu: University of Hawai'i Press.
- Crowley, Terry, John Lynch & Malcolm Ross (eds.). 2002. *The Oceanic languages*. London: Routledge.
- Culicover, Peter W. 1999. *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.
- Cunha de Oliveira, Christiane. 2005. *The language of the Apinajé people of Central Brazil*. Eugene, OR: University of Oregon dissertation.
- Curl, Tracy S. 1999. The Lakota definite articles and topic marking. *Kansas Working Papers in Linguistics* 24(2). 1–16.
- da Cruz, Aline. 2011. *Fonologia e gramática do Nheengatú [Phonology and grammar of Nheengatú]*. Utrecht: LOT.
- Danielsen, Swintha. 2007. *Baure: An Arawak language of Bolivia*. Leiden: CNWS Publications.
- Davidson, Matthew. 2002. *Studies in Southern Wakashan (Nootkan) grammar*. Buffalo, NY: University at Buffalo dissertation.
- Davis, Henry, Carry Gillon & Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90. e180–e226.
- Dayley, Jon P. 1985. *Tzutujil grammar*. Berkeley, CA: University of California Press.

- Dench, Alan Charles. 1994a. *Martuthunira. A language of the Pilbara region of Western Australia*. Canberra: Pacific Linguistics.
- Dench, Alan Charles. 1994b. The historical development of pronoun paradigms in the Pilbara region of Western Australia. *Australian Journal of Linguistics* 14(2). 155–191.
- Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization* (Typological Studies in Language 42). Amsterdam: Benjamins.
- Diessel, Holger. 2013. Distance contrasts in demonstratives. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dixon, Robert Malcolm Ward. 1980. *The languages of Australia*. Cambridge: Cambridge University Press.
- Dixon, Robert Malcolm Ward. 2002. *Australian languages: Their nature and development*. Cambridge: Cambridge University Press.
- Döhler, Christian. 2018. *A grammar of Komnzo* (Studies in Diversity Linguistics 22). Berlin: Language Science Press.
- Donnellan, Keith S. 1966. Reference and definite descriptions. *The Philosophical Review* 75(3). 281–304.
- Doornenbal, Marius. 2009. *A grammar of Bantawa*. Utrecht: LOT.
- Dorvlo, Kofi. 2008. *A grammar of Logba (Ikpana)*. Utrecht: LOT.
- Douglas, Jamie. 2018. Māori subject extraction. *Glossa* 3(1). 110.
- Dryer, Matthew S. 1989. Article-noun order. *Chicago Linguistic Society* 25. 83–97.
- Dryer, Matthew S. 1997. Are grammatical relations universal? In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type: Dedicated to T. Givón*, 115–143. Amsterdam: Benjamins.
- Dryer, Matthew S. 2013a. *A grammatical description of Kara-Lemakot*. Canberra: Asia-Pacific Linguistics.
- Dryer, Matthew S. 2013b. Definite articles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, Matthew S. 2013c. Indefinite articles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, Matthew S. 2014. Competing methods for uncovering linguistic diversity: The case of definite and indefinite articles (commentary on Davis, Gillon, and Matthewson). *Language* 90(4). 232–249.

- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dum-Tragut, Jasmine. 2009. *Armenian*. Amsterdam: Benjamins.
- Dunn, Michael John. 1999. *A grammar of Chukchi*. Canberra: Australian National University dissertation.
- Eberhard, David M. 2009. *Mamaindê grammar: A Northern Nambikwara language and its cultural context*. Utrecht: LOT.
- Ebert, Karen. 1971a. *Referenz, Sprechsituation und die bestimmten Artikel in einem Nordfriesischen Dialekt (Fering)*. Kiel: Christian-Albrechts-Universität zu Kiel dissertation.
- Ebert, Karen. 1971b. Zwei Formen des bestimmten Artikels. In Dieter Wunderlich (ed.), *Probleme und Fortschritte der Transformationsgrammatik*, 159–174. München: Hueber.
- Eilan, Naomi et al. (eds.). 2005. *Joint attention: Communication and other minds: Issues in philosophy and psychology*. Oxford: Oxford University Press.
- Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22(1). 1–25.
- Engel, Ulrich. 1988. *Deutsche Grammatik*. Heidelberg: Groos.
- Englebretson, Robert. 2003. *Searching for structure: The problem of complementation in colloquial Indonesian conversation* (Studies in Discourse and Grammar 13). Amsterdam: Benjamins.
- Enrico, John. 2003. *Haida Syntax*. Lincoln: University of Nebraska Press.
- Epstein, Richard. 1994. The development of the definite article in French. In *Perspectives on grammaticalization* (Current Issues in Linguistic Theory 109), 63–80. Amsterdam: Benjamins.
- Epstein, Richard. 2002. The definite article, accessibility, and the construction of discourse referents. *Cognitive Linguistics* 12(4). 333–378.
- Eriksson, Mats. 2009. Referring as interaction: On the interplay between linguistic and bodily practices. *Journal of Pragmatics* 41(2). 240–262.
- Erkù, Feride & Jeanette K. Gundel. 1987. The pragmatics of indirect anaphors. In *The pragmatic perspective: Selected papers from the 1985 International Pragmatics Conference*, 533–545. Amsterdam: Benjamins.
- Evans, Nicholas. 2003. *Bininj Gun-Wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics.
- Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5). 429–447.
- Everett, Daniel. 1992. *A lingua pirahã e a teoria da sintaxe: Descrição, perspectivas e teoria [The Pirahã language and a theory of syntax: Description, perspectives, and theory]*. Campinas (SP): Editora da UNICAMP.
- Everett, Daniel. 2005. Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46(4). 621–646.

- Everett, Daniel. 2009. Pirahã culture and grammar: A response to some criticisms. *Language* 85(2). 405–442.
- Farkas, Donka. 1994. Specificity and scope. In Léa Nash & Georges Tsoulas (eds.), *Actes du premier colloque langues & grammaire*, vol. 1, 119–137. Paris: Université Paris 8.
- Farkas, Donka. 2002. Specificity distinction. *Journal of Semantics* 19. 213–243.
- Fauconnier, Gilles. 1975. Pragmatic scales and logical structure. *Linguistic Inquiry* 6(3). 353–375.
- Fauconnier, Gilles. 1994. *Mental spaces*. Cambridge, MA: Cambridge University Press.
- Fauconnier, Gilles. 1997. *Mappings in thought and language*. Cambridge: Cambridge University Press.
- Fauconnier, Gilles. 2007. Mental spaces. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 351–376. Oxford: Oxford University Press.
- Fillmore, Charles J. 1975. An alternative to checklist theories of meaning. In *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 123–131. Berkeley, CA: BLS.
- Fillmore, Charles J. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences* 280(1). 20–32.
- Flick, Johanna. 2020. *Die Entwicklung des Definitartikels im Althochdeutschen: Eine kognitiv-linguistische Korpusuntersuchung* (Empirically Oriented Theoretical Morphology and Syntax 6). Berlin: Language Science Press.
- Forrester, Katerina. 2015. *The internal structure of the Mawng noun phrase*. Melbourne: University of Melbourne dissertation.
- François, Alexandre & Martine Vanhove. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In *From polysemy to semantic change: Towards a typology of lexical semantic associations*, 163–215. Amsterdam: Benjamins.
- Frank, Michael C. et al. 2008. Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108(3). 819–824.
- Fransen, Margo Astrid Eleonora. 1995. *A grammar of Limbum, a Grassfields Bantu language spoken in the north-west province of Cameroon*. Amsterdam: Vrije Universiteit te Amsterdam dissertation.
- Fraurud, Kari. 2001. Possessives with extensive use. A source of definite articles? In Irene Baron, Michael Herslund & Finn Sørensen (eds.), *Dimensions of possession*, 243–267. Amsterdam: Benjamins.
- Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100. 25–50.
- Fretheim, Thorstein & Nana Aba Appiah Amfo. 2008. Reference, determiners and descriptive content. In Henrik Høeg Müller & Alex Klinge (eds.), *Essays on nominal determination: From mor-*

- phology to discourse management* (Studies in Language Companion Series 99), 337–364. Amsterdam: Benjamins.
- Fried, Robert Wayne. 2010. *A grammar of Bao'an Tu, a Mongolic language of Northwest China*. Buffalo, NY: University of Buffalo, State University of New York dissertation.
- Frostad, Benedicte Haraldstad. 2012. *A grammar of Ughele*. Utrecht: LOT.
- Fudeman, Kirsten Anne. 1999. *Topics in the morphology and syntax of Balanta, an Atlantic language of Senegal*. Ithaca, NY: Cornell University dissertation.
- Galloway, Brent. 1993. *A grammar of Upriver Halkomelem*. Berkeley, CA: University of California Press.
- Ganenkov, Dmitry, Timur Maisak & Solmaz Merdanova. 2009. Diskursivnaja anafora v agul'skom jazyke [Discourse anaphora in Agul]. In *NeFestšrift: Stat'i v podarok (k jubileju A. E. Kibrika)*.
- Garellek, Marc & James White. 2015. Phonetics of Tongan stress. *Journal of the International Phonetic Association* 45(1). 13–34.
- Garrett, Andrew. 2010. *Basic Yurok grammar*. Berkeley, CA: Department of Linguistics University of California, Berkeley.
- Gerds, Donna B. & Thomas E. Hukari. 2004. Determiners and transitivity in Halkomelem texts. In Donna B. Gerds & Lisa Matthewson (eds.), *Studies in Salish linguistics in honor of M. Dale Kinkade*, 151–171. Missoula, MT: University of Montana.
- Gerland, Doris. 2014. Definitely not possessed? Possessive suffixes with definiteness marking function. In *Frames and concept types* (Studies in Linguistics and Philosophy 94), 269–292. Cham: Springer.
- Gernsbacher, Morton Ann & Suzanne Shroyer. 1989. The cataphoric use of the indefinite this in spoken narratives. *Memory & Cognition* 17(5). 536–540.
- Gesita, Norma & Totok Suhardijanto. 2020. Verb with suffix -nya as a definite marker in colloquial Jakartan Indonesian. In *Proceedings of the Third International Seminar on Recent Language, Literature, and Local Culture Studies, BASA, 20-21 September 2019, Surakarta, Central Java, Indonesia*, 183–195. Ghent: EAI.
- Giacon, John. 2014. *A grammar of Yuwaalaraay and Gamilaraay*. Canberra: Australian National University dissertation.
- Giannakidou, Anastasia. 1997. *The landscape of polarity items*. Groningen: University of Groningen dissertation.
- Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non)veridical dependency* (Linguistik Aktuell / Linguistics Today 23). Amsterdam: Benjamins.
- Giannakidou, Anastasia. 2017. Polarity in the semantics of natural language. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.

- Givón, Talmy. 1969. *Studies in Chibemba and Bantu grammar*. Los Angeles, CA: University of California dissertation.
- Givón, Talmy. 1978. Definiteness and referentiality. In Joseph Harold Greenberg, Charles Albert Ferguson & Edith A. Moravcsik (eds.), *Universals of human language: Syntax*, 292–330. Stanford, CA: Stanford University Press.
- Givón, Talmy. 1980. The binding hierarchy and the typology of complements. *Studies in Language* 4. 333–377.
- Givón, Talmy. 1981. On the development of the numeral ‘one’ as an indefinite marker. *Folia Linguistica Historica* 15(2.1). 35–54.
- Givón, Talmy. 1984. *Syntax: A functional-typological introduction. Volume 1*. Amsterdam: Benjamins.
- Givón, Talmy. 2011. *Ute reference grammar*. Amsterdam: Benjamins.
- Goddard, Cliff. 1985. *A grammar of Yankunytjatjara*. Alice Springs: Institute for Aboriginal Development.
- Göksel, Asli & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London: Routledge.
- Gordon, Peter. 2004. Numerical cognition without words: Evidence from Amazonia. *Science* 306(5695). 496–499.
- Graczyk, Randolph. 2007. *A grammar of Crow*. Bloomington, IN: University of Nebraska Press.
- de la Grasserie, Raoul. 1896. De l’article (morphologie et syntaxe). *Mémoires de la Société de Linguistique de Paris* 9. 285–322, 381–394.
- Green, Rebecca. 1995. *A grammar of Gurr-goni (north central Amhem Land)*. Canberra: Australian National University dissertation.
- Green, Thomas. 1999. *A lexicographic study of Ulwa*. Cambridge, MA: MIT dissertation.
- Greenberg, Joseph Harold. 1978. How does a language acquire gender markers? In Suzanne Kemmer & Keith Denning (eds.), *On language: Selected writings of Joseph H. Greenberg*, 241–270. Stanford, CA: Stanford University Press.
- Grice, H. P. 1975. Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and semantics 3: Speech acts*, 41–58. New York: Academic Press.
- Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.
- Grubic, Mira & Agata Renans. 2016. Definiteness marking on VPs/TPs in Ga and Ngamo. In Vera Hohaus & Wanda Rothe (eds.), *The Semantics of African, Asian and Austronesian Languages (TripleA)* 3, 1–15. Tübingen: Universitätsbibliothek Tübingen.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

- Hammarström, Harald & Mark Donohue. 2014. Some principles on the use of macro-areas in typological comparison. *Language Dynamics and Change* 4(1). 167–187.
- Hammarström, Harald et al. 2020. *Glottolog 4.3*. Jena: Max Planck Institute for the Science of Human History.
- Harlow, Ray. 2012. *Maori: A linguistic introduction*. Cambridge: Cambridge University Press.
- Hartmann, Dietrich. 1967. *Studien zum bestimmten Artikel in 'Morant und Galie' und anderen rheinischen Denkmälern des Mittelalters*. Gießen: Wilhelm Schmitz Verlag.
- Hartmann, Dietrich. 1982. Deixis and anaphora in German dialects: The semantics and pragmatics of two definite articles in dialectal varieties. In Jürgen Weissenborn & Wolfgang Klein (eds.), *Here and there: Cross-linguistic studies on deixis and demonstration*, 187–207. Amsterdam: Benjamins.
- Hartmann, Iren, Martin Haspelmath & Michael Cysouw. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language* 38(3). 463–484.
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Haspelmath, Martin. 1999. Explaining article-possessor complementarity: Economic motivation in noun phrase syntax. *Language* 75(2). 227–43.
- Haspelmath, Martin. 2008. Parametric versus functional explanations of syntactic universals. In Theresa Biberauer (ed.), *The limits of syntactic variation* (Linguistik Aktuell / Linguistics Today 132), 75–107. Amsterdam: Benjamins.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.
- Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm Humanities Press.
- Hawkins, John A. 1991. On (in)definite articles: Implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27(2). 405–42.
- Heath, Jeffrey. 1984. *Functional grammar of Nunggubuyu*. Canberra: Australian Institute of Aboriginal Studies.
- Heath, Jeffrey. 1986. Syntactic and lexical aspects of nonconfigurality in Nunggubuyu (Australia). *Natural Language & Linguistic Theory* 4. 375–408.
- Heath, Jeffrey. 1999. *A grammar of Koyra Chiini: The Songhay of Timbuktu* (Mouton Grammar Library 19). Berlin: De Gruyter.
- Heath, Jeffrey. 2008. *A grammar of Jamsay* (Mouton Grammar Library 45). Berlin: De Gruyter.
- Hedberg, Nancy, Emrah Görgülü & Morgan Mamani. 2009. On definiteness and specificity in Turkish and Persian. In Frédéric Mailhot (ed.), *Proceedings of the 2009 Annual Conference of the Canadian Linguistic Association*, e1–15. Ottawa: Carleton University.

- Heim, Irene. 1988. *The semantics of definite and indefinite noun phrases*. New York: Garland.
- Heim, Irene. 1991. Artikel und Definitheit. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch zeitgenössischer Forschung*, 487–535. Berlin: De Gruyter.
- Heim, Irene. 2002. File Change semantics and the familiarity theory of definiteness. In Paul Portner & Barbara H. Partee (eds.), *Formal semantics: The essential readings*, 223–248. Oxford: Blackwell.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden, MA: Blackwell.
- Heine, Bernd. 1995. Conceptual grammaticalization and prediction. In John R. Taylor & Robert E. MacLaury (eds.), *Language and the cognitive construal of the world* (Trends in Linguistics. Studies and Monographs 82), 119–136. Berlin: De Gruyter.
- Heine, Bernd. 1997. *Cognitive foundations of grammar*. Oxford: Oxford University Press.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd et al. 1995. *On indefinite articles*. Talk Presented at the International Conference on Functionalism at the University of New Mexico, Albuquerque.
- Heinrichs, Heinrich Matthias. 1954. *Studien zum bestimmten Artikel in den germanischen Sprachen*. Gießen: Wilhelm Schmitz Verlag.
- Hellenthal, Anne-Christie. 2010. *A grammar of Sheko*. Utrecht: LOT.
- Hendrick, Randall. 2005. Tongan determiners and semantic composition. *Language* 81(4). 907–926.
- Henri, Agnès. 2011. *Le suñwadia: Eléments de description d'une langue du Vanuatu*. Leuven: Peeters.
- Henry, Timothy Paul. 2012. *A pedagogical grammar of Ventureño Chumash: Implementing grammatical theory in grammar writing*. Santa Barbara, CA: University of California dissertation.
- Henschel, Elke & Harald Weydt. 2003. *Handbuch der deutschen Grammatik*. Berlin: De Gruyter.
- Heringa, Herman. 2011. *Appositional constructions*. Groningen: Rijksuniversiteit Groningen dissertation.
- von Heusinger, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19. 245–274.
- von Heusinger, Klaus. 2007. Referentially anchored indefinites. In Ileana Comorovski & Klaus von Heusinger (eds.), *Existence: Semantics and syntax* (Studies in Linguistics and Philosophy 84), 273–292. Springer.
- von Heusinger, Klaus. 2008. Verbal semantics and the diachronic development of DOM in Spanish. *Probus* 20(1). 1–31.
- von Heusinger, Klaus & Jaklin Kornfilt. 2005. The case of the direct object in Turkish: Semantics, syntax and morphology. *Turkic Languages* 9. 3–44.
- von Heusinger, Klaus & Petra B. Schumacher. 2019. Discourse prominence: Definition and application. *Journal of Pragmatics* 154. 117–127.

- van den Heuvel, Wilco. 2006. *Biak*. Utrecht: LOT.
- Hill, Jane H. 2006. *A grammar of Cupeño*. Berkeley: University of California Press.
- Himmelman, Nikolaus P. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. In Barbara A. Fox (ed.), *Studies in anaphora*, 205–254. Amsterdam: Benjamins.
- Himmelman, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Tübingen: Niemeyer.
- Himmelman, Nikolaus P. 2001. Articles. In Martin Haspelmath et al. (eds.), *Language typology and language universals. An international handbook*, vol. 1, 831–841. Berlin: De Gruyter.
- Hintikka, Jaakko. 1967. Individuals, possible worlds, and epistemic logic. *Noûs* 1(1). 33–62.
- Hoch, Ernst. 1964. *A Bemba grammar with exercises*. Ilondola: Language Centre.
- Hoeksema, Jack. 2013. On the natural history of Negative Polarity Items. *Linguistic Analysis* 38(1-2). 3–33.
- Hoffmann, Ludger. 2009. *Handbuch der deutschen Wortarten*. Berlin: De Gruyter.
- Hopper, Paul J. & Janice Martin. 1987. Structuralism and diachrony: The development of the indefinite article in English. In Anna Giacalone Ramat, Onofrio Carruba & Giuliano Bernini (eds.), *Papers from the 7th International Conference on Historical Linguistics (Current Issues in Linguistic Theory 48)*, 295–304. Amsterdam: Benjamins.
- Horn, Laurence Robert. 1972. *On the semantic properties of logical operators in English*. Los Angeles, CA: UCLA dissertation.
- Horn, Laurence Robert & Yasuhiko Kato. 2000. *Negation and polarity: Syntactic and semantic perspectives*. Oxford: Oxford University Press.
- Hualde, José. 1989. Double object constructions in Kirimi. In Robert Botne & Paul Newman (eds.), *Current approaches to African linguistics*, 179–189. Dordrecht: Foris.
- Hualde, José & Jon Ortiz de Urbina. 2003. *A grammar of Basque (Mouton Grammar Library 26)*. Berlin: De Gruyter.
- Huehnergard, John & Na'ama Pat-El. 2012. Third-person possessive suffixes as definite articles in Semitic. *Journal of Historical Linguistics* 2(1). 25–51.
- Ingason, Anton. 2016. *Realizing morphemes in the Icelandic noun phrase*. Philadelphia, PA: University of Pennsylvania dissertation.
- Ingham, Bruce. 2001. *English-Lakota dictionary*. Richmond, Surrey: Curzon Press.
- Ingham, Bruce. 2003. *Lakota*. München: Lincom Europa.
- Ionin, Tania. 2006. This is definitely specific: Specificity and definiteness in article systems. *Natural Language Semantics* 14(2). 175–234.
- Irani, Ava. 2019. On (in)definite expressions in American Sign Language. In Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages (Studies in Diversity Linguistics 25)*, 113–151. Berlin: Language Science Press.

- Jaggar, Philip J. 1985. *Factors governing the morphological coding of referents in Hausa narrative discourse*. Los Angeles, CA: University of California dissertation.
- Jaggar, Philip J. 2001. *Hausa*. Amsterdam: Benjamins.
- Jansen, Joana Worth. 2010. *A grammar of Yakima Ichishkiin/Sahaptin*. Eugene, OR: University of Oregon dissertation.
- Jany, Carmen. 2009. *Chimariko grammar. Areal and typological perspective*. Berkeley, CA: University of California Press.
- Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49(3). 501–536.
- Johnson, Heidi Anna. 2000. *A grammar of San Miguel Chimalapa Zoque*. Austin, TX: The University of Texas at Austin dissertation.
- Kamp, Hans. 1984. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Truth, interpretation and formation*, 1–42. Dordrecht: Foris.
- Kamp, Hans & Ágnes Bende-Farkas. 2019. Epistemic specificity from a communication-theoretic perspective. *Journal of Semantics* 36(1). 1–51.
- Kaplan, David. 1989. Demonstratives. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 481–563. Oxford: Oxford University Press.
- Karttunen, Lauri. 1968. In *What makes definite noun phrases definite?* Santa Monica, CA: The RAND Corporation.
- Karttunen, Lauri. 1974. Presuppositions and linguistic context. *Theoretical Linguistics* 1. 181–194.
- Karttunen, Lauri. 1976. Discourse referents. In *Syntax and semantics, vol. 7: Notes from the linguistic underground*, 363–385. New York: Academic Press.
- Kenesei, Istvan, Robert M. Vago & Anna Fenyvesi. 1998. *Hungarian*. London: Routledge.
- Khalilova, Zaira. 2009. *A grammar of Khwarshi*. Utrecht: LOT.
- Khan, Geoffrey. 2008. *The Neo-Aramaic dialect of Barwar*. Leiden: Brill.
- Kieviet, Paulus. 2017a. *A grammar of Rapa Nui* (Studies in Diversity Linguistics 12). Berlin: Language Science Press.
- Kieviet, Paulus. 2017b. Definiteness and referentiality in Rapa Nui: The interplay of determiners and demonstratives. *Oceanic Linguistics* 56(2). 305–338.
- Kimball, Geoffrey David. 1985. *A descriptive grammar of Koasati*. New Orleans, LA: Tulane University dissertation.
- Kita, Sotaro. 2003. Pointing. A foundational building block of human communication. In Sotaro Kita (ed.), *Pointing: Where language, culture, and cognition meet*, 1–8. Mahwah: Lawrence Erlbaum.
- Klamer, Marian. 2010. *A grammar of Teiwa* (Mouton Grammar Library 49). Berlin: De Gruyter.
- Koontz-Garboden, Andrew. 2009. Ulwa verb class morphology. *International Journal of American Linguistics* 4. 453–512.

- Korsah, Sampson. 2017. *Issues in Kwa syntax: Pronouns and clausal determiners*. Leipzig: Leipzig University dissertation.
- Krámský, Jirí. 1972. *The article and the concept of definiteness in language*. The Hague: Mouton.
- Krifka, Manfred & Renate Musan. 2012a. Information structure: Overview and linguistic issues. In Manfred Krifka & Renate Musan (eds.), *The expression of information structure*, 1–43. Berlin: De Gruyter.
- Krifka, Manfred & Renate Musan (eds.). 2012b. *The expression of information structure*. Berlin: De Gruyter.
- Krifka, Manfred et al. 1995. Genericity: An introduction. In Gregory N. Carlson & Francis Jeffrey Pelletier (eds.), *The generic book*, 1–124. Chicago: University of Chicago Press.
- Kripke, Saul. 1959. A completeness theorem in modal logic. *The Journal of Symbolic Logic* 24(1). 1–14.
- Kroeger, Paul. 2018. *Analyzing meaning: An introduction to semantics and pragmatics* (Textbooks in Language Sciences 5). Berlin: Language Science Press.
- Kung, Susan Smythe. 2007. *A descriptive grammar of Huehuetla Tepehua*. Austin, TX: University of Texas dissertation.
- de Lacey, Paul. 2001. *Predicate nominals and equatives in Maori*. University of Massachusetts. Unpublished Manuscript.
- Ladusaw, William Allen. 1980. *Polarity sensitivity as inherent scope relations*. New York: Garland.
- Lakoff, Robin. 1974. Remarks on “this” and “that”. *Chicago Linguistic Society* 10. 345–356.
- Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2001. Discourse in cognitive grammar. *Cognitive Linguistics* 12(2). 143–188.
- Langacker, Ronald W. 2002. Remarks on the English grounding system. In Frank Brisard (ed.), *Grounding: the epistemic footing of deixis and reference* (Cognitive Linguistics Research 21). Berlin: De Gruyter.
- Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.
- Lapesa, Rafael. 1960. Del demostrativo al artículo [From the demonstrative to the article]. *Nueva Revista de Filología Hispánica* 15(1). 23–44.
- Lazard, Gilbert. 2005. What are we typologists doing? In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic diversity and language theories* (Studies in Language Companion Series 75), 1–23. Amsterdam: Benjamins.

- Lefebvre, Claire. 1992. AGR in languages without person and number agreement: The case of the clausal determiner in Haitian and Fon. *Canadian Journal of Linguistics / Revue canadienne de linguistique* 37(2). 137–156.
- Lehmann, Christian. 2003. *Possession in Yucatec Maya*. 2nd edn. Erfurt: Seminar für Sprachwissenschaft der Universität Erfurt.
- Lehmann, Thomas. 1993. *A grammar of modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Leiss, Elisabeth. 2000. *Artikel und Aspekt: Die grammatischen Muster von Definitheit*. Berlin: De Gruyter.
- Levinson, Stephen C. 1987. Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics* 23(2). 379–434.
- Levinson, Stephen C. 2004. Deixis and pragmatic. In Laurence Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 97–121. Oxford: Blackwell.
- Lewis, David. 1970. General semantics. *Synthese* 22(1/2). 18–67.
- Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8. 339–359.
- Liljegren, Henrik. 2016. *A grammar of Palula* (Studies in Diversity Linguistics 8). Berlin: Language Science Press.
- Link, Godehard. 1998. *Algebraic semantics in language and philosophy*. Stanford, CA: CSLI Publications.
- Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4. 279–326.
- Löbner, Sebastian. 1998. Definite associative anaphora. In Simon Botley (ed.), *Approaches to discourse anaphora: Proceedings of DAARC96 - Discourse Anaphora and Resolution Colloquium*. Lancaster: Lancaster University.
- Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28(3). 279–333.
- Lorenzino, Gerardo. 1998. *The Angolar Creole Portuguese of Sao Tome: Its grammar and sociolinguistic history*. New York: CUNY dissertation.
- Louagie, Dana. 2019. *Noun phrases in Australian languages*. De Gruyter.
- Louagie, Dana & Jean-Christophe Verstraete. 2016. Noun phrase constituency in Australian languages: A typological study. *Linguistic Typology* 20(1). 25–80.
- Loughnane, Robyn. 2009. *A grammar of Oksapmin*. Melbourne: The University of Melbourne dissertation.
- Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5(1). e8559.
- Lupyan, Gary & Rick Dale. 2016. Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences* 20(9). 649–660.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

- MacDonald, Catherine Marie. 2014. *Functional projections and non-local relations in Tongan nominal phrases*. Toronto: University of Toronto dissertation.
- Manterola, Julen. 2007. *-a and bat Basque articles and recent contact theories*. University of the Basque Country. Unpublished Manuscript.
- Martin, Laura. 1977. *Positional roots in Kanjobal (Mayan)*. Gainesville, FL: University of Florida dissertation.
- Martin, Laura. 1998. Irrealis constructions in Mocho (Mayan). *Anthropological Linguistics* 40(2). 198–213.
- Marušič, Franc & Rok Žaucer. 2014. A definite article in the AP: Evidence from colloquial Slovenian. In Lilia Schürcks, Anastasia Giannakidou & Urtzi Etxeberria (eds.), *The nominal structure in Slavic and beyond* (Studies in Generative Grammar 116), 183–208. Berlin: De Gruyter.
- Mateo Toledo, Eladio. 2017. Q'anjob'al. In Judith Aissen, Nora C. England & Roberto Zavala Maldonado (eds.), *The Mayan languages*. London: Routledge.
- Matras, Yaron. 2012. *A grammar of Domari* (Mouton Grammar Library 59). Berlin: De Gruyter.
- Matushansky, Ora. 2006. Why rose is the rose: On the use of definite articles in proper names. In Olivier Bonami & Patricia Cabredo Hofherr (eds.), *Empirical issues in formal syntax and semantics 6: Papers from CSSP 2005*, 285–308. Paris: CSSP.
- McGregor, William. 1990. *A functional grammar of Gooniyandi*. Amsterdam: Benjamins.
- McGregor, William. 2011. *The Nyulnyul language of Dampier Land, Western Australia*. Canberra: Pacific Linguistics.
- McIntosh, Justin Daniel. 2011. *Grammatical sketch of Teotepic Chatino*. Austin, TX: The University of Texas at Austin MA Thesis.
- McIntosh, Justin Daniel. 2015. *Aspects of phonology and morphology of Teotepic Eastern Chatino*. Austin, TX: University of Texas at Austin dissertation.
- Merlan, Francesca. 1989. *Mangarayi*. London: Routledge.
- Michelena, Luis. 1971. Toponimia, léxico y gramática [Toponyms, lexicon and grammar]. *Fontes Linguae Vasconum* 9. 241–267.
- Mihas, Elena. 2010. *Essentials of Asheninka Perené grammar*. Milwaukee, WI: University of Wisconsin-Milwaukee dissertation.
- Miller, Mark T. 2007. *A grammar of West Coast Bajau*. Arlington, TX: University of Texas dissertation.
- Montague, Richard. 1970. Pragmatics and intensional logic. *Dialectica* 24(4). 277–302.
- Moravcsik, Edith A. 1969. Determination. *Working Papers on Language Universals* 1. 64–98.
- Morton, Ermel Joseph. 1962. *A descriptive grammar of Tongan (Polynesian)*. Bloomington, IN: Indiana University dissertation.

- De Mulder, Walter & Anne Carlier. 2011. The grammaticalization of definite articles. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press.
- Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven: Yale University Press.
- Newmark, Leonard, Peter Prifti & Philip Hubbard. 1982. *Standard Albanian: A reference grammar for students*. Stanford: Stanford University Press.
- Newmeyer, Frederick J. 2007. Linguistic typology requires crosslinguistic formal categories. *Linguistic Typology* 11(1). 133–157.
- Nikolaeva, Irina. 2003. Possessive affixes in the pragmatic structuring of the utterance: Evidence from Uralic. In Pirkko Suihkonen & Bernard Comrie (eds.), *Proceedings of the International symposium on deictic systems and quantification in languages spoken in Europe and North and Central Asia. Collection of papers*, 130–145. Leipzig / Izhevsk: MPI-EVA Leipzig, Udmurt State University.
- Noonan, Michael. 1992. *A grammar of Lango* (Mouton Grammar Library 7). Berlin: De Gruyter.
- Nordhoff, Sebastian. 2009. *A grammar of Upcountry Sri Lanka Malay*. Utrecht: LOT.
- Nordlinger, Rachel. 2014. Constituency and grammatical relations in Australian languages. In Harold Koch & Rachel Nordlinger (eds.), *The languages and linguistics of Australia: A comprehensive guide*, 215–261. Berlin: De Gruyter.
- Norwood, Susan. 1997. *Gramática de la lengua Sumu [Grammar of the Sumu language]*. Managua: CIDCA.
- O’Gorman, Tim. 2011. *Two kinds of definiteness in Lakhota*. University of Colorado at Boulder. Unpublished Manuscript.
- Obata, Kazuko. 2003. *A grammar of Bilua: A Papuan language of the Solomon Islands*. Canberra: Pacific Linguistics.
- Orkaydo, Ongaye Oda. 2013. *A grammar of Konso*. Utrecht: LOT.
- Ortmann, Albert. 2014. Definite article asymmetries and concept types: Semantic and pragmatic uniqueness. In Thmoas Gamerschlag et al. (eds.), *Frames and concept types* (Studies in Linguistics and Philosophy 94), 293–321. Cham: Springer.
- Otsuka, Yuko. 2000. *Ergativity in Tongan*. Oxford: University of Oxford dissertation.
- Ouzoubar, Erika. 1989. *L’évolution des déterminatifs dans le groupe nominal en vieux haut-allemand*. Paris: Paris Sorbonne dissertation.
- Overall, Simon E. 2007. *A grammar of Aguaruna*. Bundoora, Victoria: La Trobe University dissertation.
- Palmer, Bill. 2009. *Kokota grammar*. Honolulu: University of Hawai’i Press.
- Paperno, Denis. 2014. *Grammatical sketch of Beng*. Paris: INALCO.

- Park, Indrek. 2012. *A grammar of Hidatsa*. Bloomington, IN: Indiana University dissertation.
- Pat-El, Na'ama. 2009. The development of the Semitic definite article: A syntactic approach. *Journal of Semitic Studies* 54(1). 19–50.
- Pensalfini, Robert J. 1997. *Jingulu grammar, dictionary, and texts*. Massachusetts, MA: MIT dissertation.
- Perkins, Revere. 1992. *Deixis, grammar and culture*. Amsterdam: Benjamins.
- Pet, Willem J.A. 2011. *A grammar sketch and lexicon of Arawak (Lokono Dian)*. Dallas, TX: SIL International.
- Peterson, John M. 2011. *A grammar of Kharia: A South Munda language*. Leiden: Brill.
- Plaisir, Heleen. 2006. *A grammar of Lepcha*. Leiden: Universiteit Leiden dissertation.
- Plumley, John Martin. 1948. *An introductory Coptic grammar (Sahidic dialect)*. London: Home & Van Thal.
- Poesio, Massimo. 1994. Weak definites. In Mandy Harvey & Lynn Santelmann (eds.), *SALT IV*, 282–299. Ithaca, NY: Cornell University.
- Polinsky, Maria. 1992. Maori “he” revisited. *Oceanic Linguistics* 31(2). 229–250.
- Prince, Ellen F. 1981. On the inferencing of indefinite this NPs. In Aravind Joshi, Bonnie Webber & Ivan A. Sag (eds.), *In elements of discourse understanding*, 231–250. Cambridge: Cambridge University Press.
- Quine, Willard van Orman. 1953. Reference and modality. In *From a logical point of view*, 137–138. Cambridge, MA: Harvard University Press.
- Raymundo Gonzáles, Sonia et al. 2000. *Sk'exkixhtaqil yallay koq'anej. Variación dialectal en Q'anjobal [Dialectal variation in Q'anjobal]*. Cholsamaj: OKMA.
- de Rijk, Rudolf P. G. 2008. *Standard Basque: A progressive grammar*. Cambridge, MA: MIT Press.
- Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26(3). 287–350.
- Robinson, Laura C. 2008. *Dupaningan Agta: Grammar, vocabulary, and texts*. Honolulu: University of Hawai'i dissertation.
- Romero-Méndez, Rodrigo. 2008. *A reference grammar of Ayutla Mixe (Tukyo'm Ayuujk)*. Buffalo, NY: University of Buffalo, State University of New York dissertation.
- Rood, David S. & Allan Ross Taylor. 1996. Sketch of Lakhota, a Siouan language. In Ives Goddard & William C. Sturtevant (eds.), *Handbook of North American Indians (languages)*, vol. 17, 440–482. Washington: Smithsonian Institution.
- Ross, Malcolm. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2002. Siar. In Terry Crowley, John Lynch & Malcolm Ross (eds.), *The Oceanic languages*, 410–425. London: Routledge.

- Rowe, Karen. 2005. *Siar-Lak grammar essentials*. Ukarumpa: SIL.
- Rubin, Aaron. 2005. *Studies in Semitic grammaticalization*. Winona Lake, IN: Eisenbrauns.
- Rubin, Aaron. 2010. The development of the Amharic definite article and an Indonesian parallel. *Journal of Semitic Studies* 55(1). 103–114.
- Russell, Bertrand. 1905. On denoting. *Mind* 14. 479–493.
- Saah, Kofi K. 1994. *Studies in Akan syntax, acquisition, and sentence processing*. Bloomington, IN: Indiana University dissertation.
- Saah, Kofi K. 2010. Relative clauses in Akan. In Enoch Oladé Aboh & James Essegbey (eds.), *Topics in Kwa syntax*, 91–107. Dordrecht: Springer.
- Schapper, Antoinette. 2009. *Bunaq: A Papuan language of central Timor*. Canberra: Australian National University dissertation.
- Scheutz, Hannes. 1988. Determinantien und Definitheitsarten im Bairischen und Standarddeutschen. In Peter K. Stein (ed.), *Festschrift für Ingo Reiffenstein zum 60. Geburtstag*, 231–258. Göppingen: Kümmerle.
- Schmitt, Viola. 2006. *Hessische Relativsätze*. Vienna: University of Vienna MA Thesis.
- Schroeder, Christoph. 1999. *The Turkish nominal phrase in spoken discourse*. Wiesbaden: Harrassowitz.
- Schroeder, Christoph. 2006. Articles and article systems in some areas of Europe. In Giuliano Bernini & Marcia L. Schwartz (eds.), *Pragmatic organization of discourse in the languages of Europe* (Empirical Approaches to Language Typology 8), 545–611. Berlin: De Gruyter.
- Schultze-Berndt, Eva & Candide Simard. 2012. Constraints on noun phrase discontinuity in an Australian language: The role of prosody and information structure. *Linguistics* 50(5). 1015–1058.
- Schuster, Mauriz & Hans Schikola. 1984. *Sprachlehre der Wiener Mundart*. Wien: Österreichischer Bundesverlag.
- Schütz, Albert J. 2001. Tongan Accent. *Oceanic Linguistics* 40(2). 307–323.
- Schwarz, Florian. 2009. *Two types of definites in natural language*. Amherst, MA: University of Massachusetts Amherst dissertation.
- Schwarz, Florian. 2012. Maximality and definite plurals - experimental evidence. In Emmanuel Chemla, Vincent Homer & Grégoire Winterstein (eds.), *Proceedings of Sinn und Bedeutung 17*, 509–526. Paris: ENS.
- Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–559.
- Schwarz, Florian. 2014. How weak and how definite are weak definites? In Ana Aguilar-Guevara, Bert Le Bruyn & Joost Zwarts (eds.), *Weak referentiality* (Linguistik Aktuell / Linguistics Today 219), 213–135. Amsterdam: Benjamins.

- Schwarz, Florian. 2019. Weak and strong definite articles: Meaning and form across languages. In Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages* (Studies in Diversity Linguistics 25), 1–37. Berlin: Language Science Press.
- Serdobolskaya, Natalia, Maria Usacheva & Timofey Arkhangelskiy. 2019. Grammaticalization of possessive markers in the Beserman dialect of Udmurt. In Lars Johanson, Lidia Mazzitelli & Irina Nevskaya (eds.), *Possession in languages of Europe and North and Central Asia* (Studies in Language Companion Series 206), 291–311.
- Šereikaitė, Milena. 2019. Strong vs. weak definites: Evidence from Lithuanian adjectives. In Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages* (Studies in Diversity Linguistics 25), 83–111. Berlin: Language Science Press.
- Seyoum, Mulugeta. 2008. *A grammar of Dime*. Utrecht: LOT.
- Slobin, Dan Isaac. 1997. The origin of grammaticizable notions: Beyond the individual mind. In Dan Isaac Slobin (ed.), *The crosslinguistic study of language acquisition. Vol. 5. Expanding the contexts*, 265–323. Mahwah, NJ: Lawrence Erlbaum.
- Slobin, Dan Isaac. 2001. Form function relations: How do children find out what they are? In Melissa Bowerman & Stephen C. Levinson (eds.), *Language acquisition and conceptual development*, 406–449. Cambridge: Cambridge University Press.
- Smeets, Ineke. 2008. *A grammar of Mapuche* (Mouton Grammar Library 41). Berlin: De Gruyter.
- Sneddon, James Neil et al. 2010. *Indonesian reference grammar*. 2nd edn. Sydney: Allen & Unwin.
- Song, Jae Jung. 2001. *Linguistic typology: Morphology and syntax*. Harlow: Pearson Education.
- Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.
- Stalnaker, Robert. 1974. Pragmatic presuppositions. In Milton K. Munitz & Peter K. Unger (eds.), *Semantics and philosophy*, 197–214. New York: New York University Press.
- Stampe, David. 1976. Cardinal numeral systems. *Chicago Linguistic Society* 12. 594–609.
- Stassen, Leon. 1985. *Comparison and universal grammar*. Oxford: Blackwell.
- Steeman, Sander. 2011. *A grammar of Sandawe. A Khoisan language of Tanzania*. Utrecht: LOT.
- Stenson, Nancy. 2008. *Basic Irish. A grammar and workbook*. London: Routledge.
- Stirtz, Timothy M. 2011. *A grammar of Gaahmg. A Nilo-Saharan language of Sudan*. Utrecht: LOT.
- Strawson, P. F. 1950. On referring. *Mind* 59(235). 320–344.
- Suttles, Wayne. 2004. *Musqueam reference grammar*. Vancouver: UBC Press.
- Szczepaniak, Renata. 2011. *Grammatikalisierung im Deutschen*. 2nd edn. Tübingen: Narr.
- Szczepaniak, Renata & Johanna Flick (eds.). 2020. *Walking on the grammaticalization path of the definite article: Functional main and side roads* (Studies in Language Variation 23). Amsterdam: Benjamins.

- Tadmor, Uri. 2015. When culture grammaticalizes: The pronominal system of Onya Darat. In Rik De Busser & Randy LaPolla (eds.), *Language structure and environment: Social, cultural, and natural factors*, 77–98. Amsterdam: Benjamins.
- Terrill, Angela. 2003. *A grammar of Lavukaleve* (Mouton Grammar Library 30). Berlin: De Gruyter.
- Tomasello, Michael. 1999. *Cultural origins of human cognition*. Cambridge: Harvard University Press.
- Trask, Robert Lawrence. 2003. The noun phrase: Nouns, determiners and modifiers; pronouns and names. In José Hualde & Jon Ortiz de Urbina (eds.), *A grammar of Basque* (Mouton Grammar Library 26), 113–170. Berlin: De Gruyter.
- Trudgill, Peter. 2008. Linguistic and social typology. In *The handbook of language variation and change*, 707–728. London: Wiley.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Trudgill, Peter. 2015. Societies of intimates and linguistic complexity. In Rik De Busser & Randy LaPolla (eds.), *Language structure and environment: Social, cultural, and natural factors*, 133–148. Amsterdam: Benjamins.
- Tz’utujil Tinaamitaal. 2007. *Rub’eeyaal rutz’ijb’axiik ja tzijob’al Tz’utujil. Gramática normativa Tz’utujil [Normative grammar of Tz’utujil]*. Ciudad de Guatemala: Academia de las Lenguas Mayas de Guatemala.
- Van Valin, Robert D. 1977. *Aspects of Lakhota syntax*. Berkeley, CA: University of California dissertation.
- Vázquez Alvarez, Juan Jesús. 2011. *A grammar of Chol, a Mayan language*. Austin, TX: University of Texas dissertation.
- Viljoen, Melanie Helen. 2013. *A grammatical description of the Buwal language*. Bundoora, Victoria: La Trobe University dissertation.
- Völkel, Svenja. 2010. *Social structure, space and possession in Tongan culture and language: An ethnolinguistic study*. Amsterdam: Benjamins.
- Waite, Jeffrey. 1994. Determiner phrases in Maori. *Te Reo* 37. 55–70.
- Weber, Robert. 2003. *Morfología verbal del rapa nui, lengua polinésica de Isla de Pascua, y su función en la narración [Verbal morphology of Rapa Nui, a Polynesian language of the Isla de Pascuas, and its function in narration]*. Valparaíso: Ediciones Universitarias de Valparaíso.
- Whitehead, Carl R. 2006. *A reference grammar of Menya, an Angan language of Papua New Guinea*. Winnipeg: University of Manitoba dissertation.
- Wierzbicka, Anna. 1998. Anchoring linguistic typology in universal semantic primes. *Linguistic Typology* 2. 141–194.

- Wilde, Christopher P. 2008. *A sketch of the phonology and grammar of Rājbanshi*. Helsinki: University of Helsinki dissertation.
- Wilkins, David P. 1989. *Mparntwe Arrernte (Aranda) : Studies in the structure and semantics of grammar*. Canberra: Australian National University dissertation.
- Williams, Corinne J. 1980. *A grammar of Yuwaalaraay*. Pacific Linguistics.
- Williamson, Janis Shirley. 1984. *Studies in Lakhota grammar*. San Diego, CA: University of California dissertation.
- Wiltschko, Martina. 2002. The syntax of pronouns: Evidence from Halkomelem Salish. *Natural Language & Linguistic Theory* 20(1). 157–195.
- Wiltschko, Martina. 2013. Descriptive relative clauses in Austro-Bavarian German. *Canadian Journal of Linguistics / Revue canadienne de linguistique* 58(2). 157–189.
- Wojdak, Rachel. 2008. *The linearization of affixes: Evidence from Nuu-chah-nulth* (Studies in Natural Language and Linguistic Theory 73). Dordrecht: Springer.
- Wray, Alison & George W. Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117(3). 543–578.
- Yamada, Fumiko S. 2014. *Māori as a phrase-based language*. Mānoa: University of Hawai'i at Mānoa dissertation.
- Zariquiey Biondi, Roberto. 2011. *A grammar of Kashibo-Kakataibo*. Bundoora, Victoria: La Trobe University dissertation.
- Zavala, Roberto. 1992. *El kanjobal de San Miguel Acatán [Kanjobal of San Miguel Acatán]*. México, D.F.: Universidad Autónoma de México.
- Zimmermann, Malte. 2008. Quantification in Hausa. In Lisa Matthewson (ed.), *Quantification: A cross-linguistic perspective* (North-Holland Linguistic Series 64), 415–475. Bingley: Emerald.
- Zwarts, Frans. 1981. Negatief polaire uitdrukkingen. *Glott International* 4. 35–132.
- Zwarts, Frans. 1995. Nonverdic contexts. *Linguistic Analysis* 25. 286–312.
- Zwarts, Joost. 2014. Functional frames in the interpretation of weak nominals. In Ana Aguilar-Guevara, Bert Le Bruyn & Joost Zwarts (eds.), *Weak referentiality* (Linguistik Aktuell / Linguistics Today 219), 265–285. Amsterdam: Benjamins.
- Zymeri, Isa. 2004. *Colloquial Albanian*. London: Routledge.