Articles in the world’s languages

Laura Becker

University of Erlangen-Nuremberg

July 22, 2019
3.2.2 Specific referents .............................................. 66
3.2.3 Nonspecific referents ........................................ 68
3.2.4 Generic referents ............................................. 69
3.3 Defining the article types ..................................... 69
3.3.1 Articles from the definite domain (definite, anaphoric, recognitional) ........................................................................... 70
3.3.2 Articles of the indefinite domain (exclusive-specific, nonspecific, indefinite) ........................................................................ 71
3.3.3 Domain-crossing articles (inclusive-specific, referential) ......................................................................................... 72
3.4 The relation between article types and referent types ......................................................................................... 73
3.5 Summary .................................................................. 76

4 Articles in the definite domain ...................................... 78
4.1 Definite articles ...................................................... 78
4.1.1 Kaqchikel .......................................................... 79
4.1.2 Mokpe .............................................................. 82
4.1.3 Definite articles with spatial deictic referents ................................................................................................. 87
4.1.4 Definite articles with absolutely unique referents ................................................................................................. 90
4.1.5 Definite articles with proper nouns ................................................................................................. 92
4.2 Anaphoric articles .................................................... 94
4.2.1 Limbum ............................................................. 94
4.2.2 Akan ................................................................. 98
4.2.3 Komnzo ............................................................. 102
4.3 Recognitional articles ............................................. 104
4.4 Summary ............................................................... 106

5 Articles in the indefinite domain .................................... 108
5.1 Exclusive-specific articles ....................................... 108
5.1.1 Q’anjobal .......................................................... 109
5.1.2 Akan ................................................................. 112
5.1.3 Palula ............................................................... 113
5.2 Nonspecific articles ............................................... 115
5.2.1 Q’anjobal .......................................................... 115
5.2.2 Tongan ............................................................. 117
5.2.3 Lakota .............................................................. 120
5.2.4 The source of nonspecific articles ................................................................. 122
5.3 Indefinite articles ...................................................... 125
5.3.1 Tz’utujil ............................................................ 125
5.3.2 Carib ............................................................... 127
8.4 Article systems .......................................................... 246
  8.4.1 A typology of article systems .................................. 246
    8.4.1.1 Article systems that cover the referential space minimally ... 246
    8.4.1.2 Article systems that cover most of the referential space .... 248
    8.4.1.3 Article systems that cover the entire referential space ... 252
    8.4.1.4 The crosslinguistic distribution of article systems ........ 255
  8.4.2 Systems with functionally overlapping articles ................ 264

8.5 Summary ............................................................... 272

9 Conclusions ............................................................. 275

Appendix A Article properties (113-languages sample) .............. 282

Appendix B Examples ...................................................... 290
  B.1 Africa ................................................................. 290
  B.2 Eurasia ............................................................... 303
  B.3 Papunesia ............................................................. 315
  B.4 Australia ............................................................. 330
  B.5 North America ....................................................... 334
  B.6 South America ....................................................... 346
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Areal coverage of the sample</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>Areal distribution of definite and indefinite articles (Dryer 2013a)</td>
<td>9</td>
</tr>
<tr>
<td>3.1</td>
<td>Discourse situation and discourse universe</td>
<td>44</td>
</tr>
<tr>
<td>3.2</td>
<td>Structure of a referring expression</td>
<td>48</td>
</tr>
<tr>
<td>3.3</td>
<td>Link between the referents of the kinds ‘book’ and ‘author’</td>
<td>62</td>
</tr>
<tr>
<td>3.4</td>
<td>Link between the referents of the kinds ‘village’ and ‘church’</td>
<td>64</td>
</tr>
<tr>
<td>3.5</td>
<td>Anchoring of a specific referent</td>
<td>67</td>
</tr>
<tr>
<td>3.6</td>
<td>Nonspecific referent with its concept</td>
<td>68</td>
</tr>
<tr>
<td>3.7</td>
<td>Overview of referent types and article types</td>
<td>73</td>
</tr>
<tr>
<td>3.8</td>
<td>Hierarchy of the major referent types</td>
<td>74</td>
</tr>
<tr>
<td>7.1</td>
<td>Grammaticalization path from demonstrative to general noun marker</td>
<td>165</td>
</tr>
<tr>
<td>7.2</td>
<td>Cut-off point between demonstratives and articles</td>
<td>166</td>
</tr>
<tr>
<td>7.3</td>
<td>Demonstrative with anaphoric article</td>
<td>168</td>
</tr>
<tr>
<td>7.4</td>
<td>Demonstrative with deictic and anaphoric functions</td>
<td>169</td>
</tr>
<tr>
<td>7.5</td>
<td>Demonstrative with uniqueness-based definite article</td>
<td>170</td>
</tr>
<tr>
<td>7.6</td>
<td>Demonstrative with definite article</td>
<td>171</td>
</tr>
<tr>
<td>7.7</td>
<td>Grammaticalization path from the numeral ‘one’ to a nominal marker</td>
<td>174</td>
</tr>
<tr>
<td>7.8</td>
<td>Cut-off point between the numeral ‘one’ and indefinite articles</td>
<td>176</td>
</tr>
<tr>
<td>8.1</td>
<td>Article inflection in the six macro areas</td>
<td>198</td>
</tr>
<tr>
<td>8.2</td>
<td>Posterior distributions for article inflection across the six macro areas</td>
<td>199</td>
</tr>
<tr>
<td>8.3</td>
<td>Categories of article inflection</td>
<td>200</td>
</tr>
<tr>
<td>8.4</td>
<td>Categories of article inflection in the six macro areas</td>
<td>201</td>
</tr>
<tr>
<td>8.5</td>
<td>Inflectional behaviour of following and preceding articles</td>
<td>202</td>
</tr>
<tr>
<td>8.6</td>
<td>Posterior distributions for article inflection depending on the position</td>
<td>203</td>
</tr>
<tr>
<td>8.7</td>
<td>Inflectional behaviour of different article types</td>
<td>203</td>
</tr>
<tr>
<td>8.8</td>
<td>Posterior distributions for article inflection depending on the article type</td>
<td>204</td>
</tr>
</tbody>
</table>
## List of Tables

1.1 Summary of referent types and article types ........................................... 5  
1.2 Primary data collection ............................................................................. 11  
1.3 Languages from the macroarea “Africa” ....................................................... 12  
1.4 Languages from the macroarea “Eurasia” ...................................................... 13  
1.5 Languages from the macroarea “Papunesia” ............................................... 14  
1.6 Languages from the macroarea “Australia” ................................................... 15  
1.7 Languages from the macroarea “North America” ......................................... 16  
1.8 Languages from the macroarea “South America” ......................................... 17  
3.1 Traditional view of specificity and nonspecificity ........................................... 66  
3.2 Summary of referent types and article types ................................................. 77  
4.1 Mokpe article paradigm ............................................................................. 83  
4.2 Kaqchikel demonstratives .......................................................................... 89  
4.3 Limbum demonstratives ............................................................................ 94  
5.1 Plural exponents of the exclusive-specific article in Q’anjobal ...................... 109  
5.2 Tongan article paradigm ........................................................................... 117  
5.3 Paradigm of Tongan adnominal possessive markers (A-set) ....................... 119  
5.4 Paradigm of Tongan adnominal possessive markers (O-set) ....................... 119  
5.5 Lakota specific and nonspecific article paradigm ......................................... 120  
6.1 Bemba noun class markers ....................................................................... 138  
6.2 Halkomelem article exponents .................................................................. 150  
6.3 The system of determiners in Baure as presented in Danielsen (2007: 311) ... 154  
7.1 Macedonian definite article paradigm ......................................................... 172  
7.2 Characteristic properties of indefinite articles (Heine 1997: 68) .................... 175  
7.3 Relevant contexts for adnominal NPIs ......................................................... 180  
8.1 German article paradigm .......................................................................... 187
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.38 Distribution of article systems in the sample (third series)</td>
<td>253</td>
</tr>
<tr>
<td>8.39 Article exponents in Ayoreo (Ciucci 2016: 456-462, 475-481)</td>
<td>254</td>
</tr>
<tr>
<td>8.40 Distribution of article systems in the reduced sample</td>
<td>257</td>
</tr>
<tr>
<td>8.41 Observed proportions of the first system series</td>
<td>260</td>
</tr>
<tr>
<td>8.42 Observed proportions of the second system series</td>
<td>261</td>
</tr>
<tr>
<td>8.43 Observed proportions of the third system series</td>
<td>263</td>
</tr>
<tr>
<td>8.44 Systems with overlapping articles in the definite domain</td>
<td>265</td>
</tr>
<tr>
<td>8.45 Systems with domain crossing and overlapping articles</td>
<td>267</td>
</tr>
<tr>
<td>9.1 First series</td>
<td>278</td>
</tr>
<tr>
<td>9.2 Second series</td>
<td>278</td>
</tr>
<tr>
<td>9.3 Third series</td>
<td>278</td>
</tr>
<tr>
<td>A.1 Article properties (Africa)</td>
<td>283</td>
</tr>
<tr>
<td>A.2 Article properties (Eurasia)</td>
<td>284</td>
</tr>
<tr>
<td>A.3 Article properties (Papunesia)</td>
<td>285</td>
</tr>
<tr>
<td>A.4 Article properties (Australia)</td>
<td>286</td>
</tr>
<tr>
<td>A.5 Article properties (North America)</td>
<td>287</td>
</tr>
<tr>
<td>A.6 Article properties (South America)</td>
<td>288</td>
</tr>
<tr>
<td>A.7 Inflectional behaviour of different articles within single languages</td>
<td>289</td>
</tr>
</tbody>
</table>
### Abbreviations

<table>
<thead>
<tr>
<th>1,2,3</th>
<th>1st, 2nd, 3rd person</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL</td>
<td>ablative</td>
</tr>
<tr>
<td>ABS</td>
<td>absolutive</td>
</tr>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>ACTL</td>
<td>actual</td>
</tr>
<tr>
<td>ACT</td>
<td>active</td>
</tr>
<tr>
<td>ADDIT</td>
<td>additive</td>
</tr>
<tr>
<td>ADDR</td>
<td>addressee marker</td>
</tr>
<tr>
<td>ADJ</td>
<td>adjective, adjectivizer</td>
</tr>
<tr>
<td>ADMON</td>
<td>admonition</td>
</tr>
<tr>
<td>ADNOM</td>
<td>adnominal</td>
</tr>
<tr>
<td>ADVERS</td>
<td>adversative</td>
</tr>
<tr>
<td>ADV</td>
<td>adverbial</td>
</tr>
<tr>
<td>AFFECT</td>
<td>affective</td>
</tr>
<tr>
<td>AFF</td>
<td>affirmative</td>
</tr>
<tr>
<td>AG</td>
<td>agentive, agent</td>
</tr>
<tr>
<td>ALL</td>
<td>allative</td>
</tr>
<tr>
<td>AL</td>
<td>alienable</td>
</tr>
<tr>
<td>AM</td>
<td>association marker</td>
</tr>
<tr>
<td>ANA</td>
<td>anaphoric</td>
</tr>
<tr>
<td>ANT</td>
<td>anterior</td>
</tr>
<tr>
<td>AN</td>
<td>animate</td>
</tr>
<tr>
<td>AOR</td>
<td>aorist</td>
</tr>
<tr>
<td>APPL</td>
<td>applicative</td>
</tr>
<tr>
<td>APUD</td>
<td>apudessive</td>
</tr>
<tr>
<td>ART</td>
<td>article</td>
</tr>
<tr>
<td>ASSERT</td>
<td>assertive</td>
</tr>
<tr>
<td>ATTR</td>
<td>attributive marker</td>
</tr>
<tr>
<td>AUGM</td>
<td>augmentative</td>
</tr>
<tr>
<td>AUG</td>
<td>augmented number</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary verb</td>
</tr>
<tr>
<td>BEN</td>
<td>benefactive</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>CD</td>
<td>clausal determiner</td>
</tr>
<tr>
<td>CERT</td>
<td>certainty</td>
</tr>
<tr>
<td>CIRC</td>
<td>circumstantial</td>
</tr>
<tr>
<td>CJ</td>
<td>conjoint tense</td>
</tr>
<tr>
<td>CLEFT</td>
<td>cleft marker</td>
</tr>
<tr>
<td>CL</td>
<td>class marker, classifier</td>
</tr>
<tr>
<td>CNTG</td>
<td>contiguous</td>
</tr>
<tr>
<td>COMM</td>
<td>common noun</td>
</tr>
<tr>
<td>COMPAR</td>
<td>comparative</td>
</tr>
<tr>
<td>COMPL</td>
<td>completive</td>
</tr>
<tr>
<td>COMP</td>
<td>complementizer</td>
</tr>
<tr>
<td>COM</td>
<td>comitative</td>
</tr>
<tr>
<td>COND</td>
<td>conditional</td>
</tr>
<tr>
<td>CONF</td>
<td>confirmation marker</td>
</tr>
<tr>
<td>CONJ</td>
<td>conjunction</td>
</tr>
<tr>
<td>CONSEQ</td>
<td>consequential</td>
</tr>
<tr>
<td>CONS</td>
<td>consecutive</td>
</tr>
<tr>
<td>CONT</td>
<td>continuous/continuative</td>
</tr>
<tr>
<td>COORD</td>
<td>coordinator</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>COS</td>
<td>change of state</td>
</tr>
<tr>
<td>CUSTOM</td>
<td>customary verbal suffix</td>
</tr>
<tr>
<td>CV</td>
<td>converb</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DEB</td>
<td>debitive</td>
</tr>
<tr>
<td>DECAUS</td>
<td>decausative</td>
</tr>
<tr>
<td>DECL</td>
<td>declarative</td>
</tr>
<tr>
<td>DEF</td>
<td>definite</td>
</tr>
<tr>
<td>DEM</td>
<td>demonstrative</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>DEP</td>
<td>dependent aspect</td>
</tr>
<tr>
<td>DESID</td>
<td>desiderative</td>
</tr>
<tr>
<td>DEV</td>
<td>devaluation</td>
</tr>
<tr>
<td>DIM</td>
<td>diminutive</td>
</tr>
<tr>
<td>DIR</td>
<td>directional</td>
</tr>
<tr>
<td>DIST</td>
<td>distal</td>
</tr>
<tr>
<td>DJ</td>
<td>disjoint tense</td>
</tr>
<tr>
<td>DS</td>
<td>different subject (switch reference)</td>
</tr>
<tr>
<td>DUR</td>
<td>durative</td>
</tr>
<tr>
<td>DU</td>
<td>dual</td>
</tr>
<tr>
<td>DWELL</td>
<td>dweller</td>
</tr>
<tr>
<td>EMPH</td>
<td>emphasis</td>
</tr>
<tr>
<td>EPIST</td>
<td>epistemic possibility</td>
</tr>
<tr>
<td>ERG</td>
<td>ergative</td>
</tr>
<tr>
<td>EVID</td>
<td>evidential</td>
</tr>
<tr>
<td>EV</td>
<td>eventive</td>
</tr>
<tr>
<td>EXCL</td>
<td>exclamative</td>
</tr>
<tr>
<td>EXIST</td>
<td>existential copula</td>
</tr>
<tr>
<td>EXSPEC</td>
<td>exclusive-specific</td>
</tr>
<tr>
<td>EX</td>
<td>exclusive</td>
</tr>
<tr>
<td>FACT</td>
<td>factitive</td>
</tr>
<tr>
<td>FOC</td>
<td>focus</td>
</tr>
<tr>
<td>FRAME</td>
<td>framing, background</td>
</tr>
<tr>
<td>FRUST</td>
<td>frustrative</td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
</tr>
<tr>
<td>F</td>
<td>feminine</td>
</tr>
<tr>
<td>GENPOSS</td>
<td>general possessor</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>GOAL</td>
<td>goal</td>
</tr>
<tr>
<td>HAB</td>
<td>habitual</td>
</tr>
<tr>
<td>HESIT</td>
<td>hesitation</td>
</tr>
<tr>
<td>HODFUT</td>
<td>hodiernal (today’s) future</td>
</tr>
<tr>
<td>HODPST</td>
<td>hodiernal (today’s) past</td>
</tr>
<tr>
<td>HON</td>
<td>honorific</td>
</tr>
<tr>
<td>HORT</td>
<td>hortative</td>
</tr>
<tr>
<td>HUM</td>
<td>human</td>
</tr>
<tr>
<td>IDEF</td>
<td>ideophone</td>
</tr>
<tr>
<td>IMM</td>
<td>immediate</td>
</tr>
<tr>
<td>IMPERS</td>
<td>impersonal</td>
</tr>
<tr>
<td>IMPL</td>
<td>implicated</td>
</tr>
<tr>
<td>IMP</td>
<td>imperative</td>
</tr>
<tr>
<td>INAL</td>
<td>inalienable</td>
</tr>
<tr>
<td>INAN</td>
<td>inanimate</td>
</tr>
<tr>
<td>INCH</td>
<td>inchoative</td>
</tr>
<tr>
<td>INDEF</td>
<td>indefinite</td>
</tr>
<tr>
<td>IND</td>
<td>indicative</td>
</tr>
<tr>
<td>INESS</td>
<td>inessive</td>
</tr>
<tr>
<td>INFER</td>
<td>inferential</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>INSPEC</td>
<td>inclusive-specific</td>
</tr>
<tr>
<td>INSTR</td>
<td>instrumental</td>
</tr>
<tr>
<td>INTR</td>
<td>intransitive</td>
</tr>
<tr>
<td>INT</td>
<td>intensive</td>
</tr>
<tr>
<td>IN</td>
<td>inclusive</td>
</tr>
<tr>
<td>IRR</td>
<td>irrealis</td>
</tr>
<tr>
<td>ITER</td>
<td>iterative</td>
</tr>
<tr>
<td>LAT</td>
<td>lative</td>
</tr>
<tr>
<td>LNK</td>
<td>linker</td>
</tr>
<tr>
<td>MED</td>
<td>medial</td>
</tr>
<tr>
<td>MID</td>
<td>middle</td>
</tr>
<tr>
<td>MIN</td>
<td>minimal number</td>
</tr>
<tr>
<td>M</td>
<td>masculine</td>
</tr>
<tr>
<td>NARR</td>
<td>narrative marker, narrative register</td>
</tr>
<tr>
<td>NEARFUT</td>
<td>near future</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>NFUT</td>
<td>nonfuture</td>
</tr>
<tr>
<td>NF</td>
<td>non-feminine</td>
</tr>
<tr>
<td>NHUM</td>
<td>non-human</td>
</tr>
<tr>
<td>NMLZ</td>
<td>nominalizer</td>
</tr>
<tr>
<td>NMOD</td>
<td>noun modification marker</td>
</tr>
<tr>
<td>NM</td>
<td>non-masculine</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>NOST</td>
<td>discourse particle marking nostalgia</td>
</tr>
<tr>
<td>NPI</td>
<td>negative polarity item</td>
</tr>
<tr>
<td>NPOSS</td>
<td>nonpossessed</td>
</tr>
<tr>
<td>NPST</td>
<td>nonpast</td>
</tr>
<tr>
<td>NSG</td>
<td>non-singular</td>
</tr>
<tr>
<td>NSPEC</td>
<td>nonspecific</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>NTR</td>
<td>neutral aspect</td>
</tr>
<tr>
<td>NUM</td>
<td>numeral marker</td>
</tr>
<tr>
<td>NVIS</td>
<td>non-visible</td>
</tr>
<tr>
<td>N</td>
<td>neuter</td>
</tr>
<tr>
<td>OBLG</td>
<td>obligation marker</td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
</tr>
<tr>
<td>OBV</td>
<td>obviative</td>
</tr>
<tr>
<td>ORIG</td>
<td>origin (spatial value)</td>
</tr>
<tr>
<td>O</td>
<td>object marker</td>
</tr>
<tr>
<td>PART</td>
<td>partitive</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PC</td>
<td>paucal</td>
</tr>
<tr>
<td>PERF</td>
<td>perfect</td>
</tr>
<tr>
<td>PERSIST</td>
<td>perfective</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>PM</td>
<td>phrasal marker</td>
</tr>
<tr>
<td>POLIT</td>
<td>politeness marker</td>
</tr>
<tr>
<td>POL</td>
<td>polarity</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>POST</td>
<td>postposition</td>
</tr>
<tr>
<td>POT</td>
<td>potential</td>
</tr>
<tr>
<td>PRED</td>
<td>predicative</td>
</tr>
<tr>
<td>PREP</td>
<td>preposition</td>
</tr>
<tr>
<td>PRESUMP</td>
<td>presumptive</td>
</tr>
<tr>
<td>PRES</td>
<td>presentational</td>
</tr>
<tr>
<td>PRIV</td>
<td>private</td>
</tr>
<tr>
<td>PROG</td>
<td>progressive</td>
</tr>
<tr>
<td>PROM</td>
<td>prominence marker</td>
</tr>
<tr>
<td>PROP</td>
<td>proper noun marker</td>
</tr>
<tr>
<td>PROX</td>
<td>proximal</td>
</tr>
<tr>
<td>PRS</td>
<td>present</td>
</tr>
<tr>
<td>PST1</td>
<td>past 1, recent past</td>
</tr>
<tr>
<td>PST2</td>
<td>past2, intermediate past</td>
</tr>
<tr>
<td>PST3</td>
<td>past 3, remote past</td>
</tr>
<tr>
<td>PST:UW</td>
<td>unwitnessed past</td>
</tr>
<tr>
<td>PST:W</td>
<td>witnessed past</td>
</tr>
<tr>
<td>PAST</td>
<td>past</td>
</tr>
<tr>
<td>PTCP</td>
<td>participle</td>
</tr>
<tr>
<td>PURP</td>
<td>purposive</td>
</tr>
<tr>
<td>P</td>
<td>patient</td>
</tr>
<tr>
<td>QUOT</td>
<td>quotative</td>
</tr>
<tr>
<td>Q</td>
<td>question/interrogative marker</td>
</tr>
<tr>
<td>RC</td>
<td>relative clause marker</td>
</tr>
<tr>
<td>REAL</td>
<td>realis</td>
</tr>
<tr>
<td>RECIP</td>
<td>reciprocal</td>
</tr>
<tr>
<td>RECOG</td>
<td>recognitional</td>
</tr>
<tr>
<td>RECPST</td>
<td>recent past</td>
</tr>
<tr>
<td>RED</td>
<td>reduplication</td>
</tr>
<tr>
<td>REFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>REF</td>
<td>referential</td>
</tr>
<tr>
<td>RELAT</td>
<td>relational</td>
</tr>
<tr>
<td>REL</td>
<td>relativizer</td>
</tr>
<tr>
<td>REMFUT</td>
<td>remote future</td>
</tr>
<tr>
<td>REMPST</td>
<td>remote past</td>
</tr>
<tr>
<td>REM</td>
<td>remote</td>
</tr>
<tr>
<td>REPORT</td>
<td>reported</td>
</tr>
<tr>
<td>REP</td>
<td>repetitive</td>
</tr>
<tr>
<td>RES</td>
<td>resultative</td>
</tr>
<tr>
<td>SEQ</td>
<td>sequential</td>
</tr>
<tr>
<td>SER</td>
<td>verb serializer</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SIM</td>
<td>simultaneous</td>
</tr>
<tr>
<td>SPC</td>
<td>nominal specifier</td>
</tr>
<tr>
<td>SRC</td>
<td>source</td>
</tr>
<tr>
<td>SS</td>
<td>same subject (switch reference)</td>
</tr>
<tr>
<td>STAT</td>
<td>stative</td>
</tr>
<tr>
<td>STI</td>
<td>indirect stance marker</td>
</tr>
<tr>
<td>SUBJ</td>
<td>subjunctive</td>
</tr>
<tr>
<td>SUB</td>
<td>subordinate, subordinator</td>
</tr>
<tr>
<td>SUPER</td>
<td>super localization</td>
</tr>
<tr>
<td>SUP</td>
<td>superlative</td>
</tr>
<tr>
<td>S</td>
<td>subject</td>
</tr>
<tr>
<td>TC</td>
<td>topic change marker</td>
</tr>
<tr>
<td>TEMP</td>
<td>temporal marker</td>
</tr>
<tr>
<td>THEME</td>
<td>theme (voice)</td>
</tr>
<tr>
<td>TOP</td>
<td>topic</td>
</tr>
<tr>
<td>TRI</td>
<td>trial</td>
</tr>
<tr>
<td>TR</td>
<td>transitive</td>
</tr>
<tr>
<td>VBLZ</td>
<td>verbalizer</td>
</tr>
<tr>
<td>VENT</td>
<td>ventive</td>
</tr>
<tr>
<td>VICIN</td>
<td>vicinity</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>VIS</td>
<td>visible</td>
</tr>
<tr>
<td>A</td>
<td>agent argument in transitive clauses</td>
</tr>
</tbody>
</table>

xiii
Chapter 1

Introduction

1.1 Background and objectives

At first sight, linguists seem to have an intuitive understanding of what articles are as a category: they usually occur with nouns and mark them as definite or indefinite. However, a closer look at how the label of articles is used in the linguistic literature reveals that we still lack a proper understanding of what articles are.

There are three main problems: First, articles are often classified as definite or indefinite even if their functions differ considerably from what we would usually call definite or indefinite articles; in many cases 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 define articles and also 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. It is often their similarity to English articles that is taken as a criterion to treat the marker as an article or not. 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 referred to as proper or personal article.

What these three different uses of the article as a label show is that it is still not clear which of its properties in well-known languages are relevant to define articles language-independently. We still lack a precise definition of their function, which is why markers are sometimes treated as articles or not based on the functions of articles in e.g. English. In addition, a principled distinction and definition of different article types are still not available, which is why we see the labels of definite and indefinite articles applied to different markers across languages with a wide range of functions. What complicates the picture even more is that their occurrence with nouns is sometimes taken as sole criterion to treat markers as articles, leading to an even less coherent
use of the notion. Hence, even though most linguists may have an intuition about articles, we still do not have principled criteria that can be applied to any language in order to decide whether a given marker is an article or not, and if so, which type of article it is.

This is surprising especially given the large amount of literature that has dealt with articles in single languages and language families. Since articles can be approached from various point of views, for instance, with respect to their semantic and/or syntactic properties, it is impossible to provide an appropriate overview of this literature here. The main point is that while all these approaches may have suitable criteria to discuss articles within single languages or language families, there still is no good way of comparing these results against each other, since the notion of articles is understood in so many different ways. This study aims at closing this gap by defining articles as a crosslinguistic category and analyzing the crosslinguistic trends of their properties.

During 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. Articles, on the other hand, have not yet been discussed in detail as a crosslinguistic category. However, even though they have not been the main point of attention in a typological study, a number of typologically oriented works that discuss articles have to be mentioned.

The studies that probably come closest to a detailed typological discussion of articles are: Dryer (1989, 2013a,b, 2014), Ghomeshi et al. (2009), Himmelmann (1997), Leiss (2000), Schroeder (2006). Ghomeshi et al. (2009) is a collection of studies on determiners from a variety of languages, but articles are only one of various phenomena discussed. Both Schroeder (2006) and Leiss (2000) offer a comparative studies on articles, focusing on articles in Europe. While Schroeder (2006) provides an overview of different properties of the articles found in various areas of Europe, Leiss (2000) investigates the relation between aspect and articles in a number of Indo-European languages, arguing that articles emerged as a consequence of the loss of aspeptual distinctions in the verbal domain.

Dryer (1989, 2013a,b), on the other hand, discusses definite and indefinite articles from a broader crosslinguistic perspective. Dryer (1989) investigates the correlation between the order of articles and nouns with VO and OV word orders. Therefore, his focus does not lie on articles as such, even though it is based on probably the largest crosslinguistic sample of articles available until now. Similarly, Dryer (2013a,b) 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 at a large crosslinguistic scale, these two chapters can only offer 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. Dryer (2014) is a reply to
criticisms of these two WALS chapters and clarifies a number of methodological questions especially concerning the definition of articles. It is therefore also a condensed discussion of articles, crosslinguistic trends with respect to different article types, and selected peculiarities of articles. Himmelmann (1997) offers another crosslinguistic study discussing the diachronic development of nominal expression and demonstratives to lead to different types of fixed nominal syntagmas including definite articles. Since it is concerned with the definite domain only, different types of articles in the indefinite domain are not part of that study.

Thus, to the best of my knowledge, a detailed and systematic crosslinguistic study of articles is still not available. I propose a language-independent definition of articles as a crosslinguistic category based on three main criteria: their domain (the noun phrase), their function (expressing referentiality), and their systematic occurrence. I define different referential functions, called referent types, which are used to define different types of articles. On the basis of a sample including 113 languages with 148 articles, I show detailed examples from different languages of the world for the 8 major article types that can be distinguished based on their functions. Having defined and illustrated different types of articles, this study discusses the differences between certain article types and other elements, such as definite articles and demonstratives. Then I explore the crosslinguistic variation and trends of other properties that articles have, beginning with article inflection and article paradigms, and then discussing crosslinguistic trends concerning different article types and article systems.

The focus of this study is a crosslinguistic comparison of articles by capturing the variation and trends. Therefore, it does not aim at an in-depth analysis of articles or article systems of particular languages: even though I provide a number of case studies, 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) in the languages in question.

1.2 Overview

Chapter 2: Articles as a crosslinguistic category In Chapter 2, I define articles as markers that occur systematically with nouns to mark the referential function of the noun. This definition includes three types of criteria: the function, domain, and distribution criteria. The function criterion allows us to distinguish between the main types of articles on the basis of their referential functions: anaphoric, recognitional, definite, inclusive-specific, exclusive-specific, indefinite, nonspecific, and referential articles. The form criterion ensures that only markers which occur in the nominal domain 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. Systematically does not
mean obligatory 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. This holds in either direction: articles can be absent in contexts in which their presence would be required, but they can also be present in contexts in which they are not required to indicate a certain referential value. 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: Defining different types of articles  Chapter 3 deals with referentiality and different referent types which are used as a basis to define different types of articles. In order to define the referent types, a number of important notions related to discourse and referentiality are introduced in section 3.1. Then section 3.2 defines different referent types to distinguish forms (nouns) from their meanings (referents). I use the notion of identifiability as a basis for the distinction of different referent types. Identifiability is defined as the discourse participants’ knowledge and commitment to this knowledge about the referent. It can be used to distinguish a number of different types of definite referents, as well as specific and nonspecific referents. The last part of this chapter defines different types of articles according to which (combinations of) referent types they express. On this basis, the following 8 article types are defined in section 3.3: definite (def), anaphoric (ana), recognitional (rec), exclusive-specific (exspec), nonspecific (nspec), indefinite (indef), inclusive-specific (inspec), and referential (ref) articles. An overview summarizing the relation between article and referent types is provided in Table 1.1. The referent types are listed and briefly described in the different rows of Table 1.1, while the article types are represented to the left of the Table with arrows indicating their functional domains.
Table 1.1: Summary of referent types and article types

<table>
<thead>
<tr>
<th>Referent Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{dei}$</td>
<td>A deictic referent $R_{dei}$ is identifiable based on its unambiguous link to an object in the discourse situation.</td>
</tr>
<tr>
<td>$R_{recog}$</td>
<td>A recognitional referent $R_{recog}$ is identifiable based on shared previous experience or common knowledge.</td>
</tr>
<tr>
<td>$R_{abs.u}$</td>
<td>An absolutely unique referent $R_{abs.u}$ is identifiable because it is the only referent of its kind.</td>
</tr>
<tr>
<td>$R_{ana}$</td>
<td>An anaphoric referent $R_{ana}$ is identifiable based on shared identity with a previously mentioned referent.</td>
</tr>
<tr>
<td>$R_{brid}$</td>
<td>A bridging referent $R_{brid}$ is identifiable based on its unambiguous link to another previously mentioned referent.</td>
</tr>
<tr>
<td>$R_{cont.u}$</td>
<td>A contextually unique referent $R_{cont.u}$ is identifiable because it is the only salient referent of its kind in the discourse situation.</td>
</tr>
<tr>
<td>$R_{est}$</td>
<td>An establishing referent $R_{est}$ is marked as identifiable and thus constructed as such.</td>
</tr>
<tr>
<td>$R_{spec}$</td>
<td>A specific referent $R_{spec}$ is not identifiable but linked to a particular referent of its kind.</td>
</tr>
<tr>
<td>$R_{nspec}$</td>
<td>A nonspecific referent $R_{nspec}$ corresponds to a single, but no particular (any), referent of its kind.</td>
</tr>
<tr>
<td>$R_{gen}$</td>
<td>A generic referent $R_{gen}$ corresponds to the set of all referents of its kind.</td>
</tr>
</tbody>
</table>

Chapter 4: Articles in the definite domain  In Chapter 4, I discuss examples of the three article types in the definite domain: definite, anaphoric, and recognitional articles. For definite articles, I present two examples from Kaq’chikel (Maya, Guatemala) and Mokpe (Bantu, Cameroon) in detail in section 4.1. Then I discuss the expression of deictic and absolutely unique referents by definite articles, concluding that they can but do not need be expressed by definite articles. Section 4.2 deals with anaphoric articles on the basis of examples from Limbum (Bantu, Cameroon), Akan (Kwa, Ghana), and Komnzo (Morehead-Wasur, Papua New Guinea). Finally, section 4.3 is concerned with recognitional articles, mostly found in the languages in Papunesia and Australia.
Chapter 5: Articles in the indefinite domain  There are three types of articles in the indefinite domain: exclusive-specific, nonspecific, and indefinite articles. This chapter presents examples for each of these three article types, with examples from Q’anjobal (Maya, Guatemala), Akan (Kwa, Ghana), and Palula (Dardic, Pakistan) for exclusive-specific articles in section 5.1. Then section 5.2 introduces nonspecific articles in Q’anjobal, Tongan (Oceanic, Tonga), and Lakota (Sioux, USA). Finally, section 5.3 discusses indefinite articles with examples from Tz’utujil (Maya, Guatemala), Carib (Cariban, Suriname), and Bonan (Mongolic, China). The second part of this section is concerned with a subtype of indefinite articles, called presentational articles. Their use is restricted to discourse prominent referents, which almost always correlates with specific referents in the indefinite domain. However, I show that this is not necessarily the case: presentational articles are not restricted to specific referents but can occur with nonspecific referents as well.

Chapter 6: Domain-crossing articles  Two types of articles express referents of both the definite and indefinite domain, namely inclusive specific and referential articles, which are presented in more detail in this chapter. In the first section, I discuss inclusive specific articles with examples from Bemba (Bantu, Zambia), Tongan (Oceanic, Tonga), and Tepehua (Totonacan, Mexico). The second section is concerned with referential articles, presenting their properties in detail in Rapa Nui (Oceanic, Chile), Halkomelem (Halkomelem, Salishan), and Baure (Arawakan, Bolivia).

Chapter 7: The distinction between articles and similar categories  Chapter 7 delimits certain article types from other similar categories, namely demonstratives, the numeral ‘one’, and negative polarity items. Demonstratives and the numeral ‘one’ are difficult to distinguish from definite and indefinite articles because of the diachronic relation of these elements: definite articles often originate from demonstratives, as well as indefinite articles do so from the numeral ‘one’. In section 7.1, I define the cut-off point between (adnominal) demonstratives and articles in the following way: markers that encode exophoric deixis only or extend to endophoric deixis are treated as demonstratives. Markers that encode endophoric deixis (discourse anaphora) only or extend to uniqueness are viewed as articles. Also, markers that encode all three functions of exophoric and endophoric deixis as well as uniqueness-based identifiability are treated as articles.

Accordingly, section 7.2 argues how the distinction between the numeral ‘one’ and indefinite and specific articles can be based on the scale from the functions of individuation, presentative marking, specific marking, and nonspecific marking. Although this scale is usually taken to describe the diachronic development of indefinite articles, it can be used to delimit the numeral from articles: I argue that 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 is not systematic, the marker is not counted as an article. If the marker is used consistently to signal the non-
identifiability of the referent, even though it is restricted by discourse prominence or information structure, it is treated as an article (presentational article).

Another article type, nonspecific articles, also needs to be delimited from a subgroup of elements usually called NPIs (negative polarity items). I argue in section 7.3 that even though nonspecific articles and certain NPIs have the same referential functions, NPIs lead to pragmatics effects that nonspecific articles do not have, which manifests itself in the optionality of NPIs in the sense that 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; they are used systematically in nonspecific contexts.

Chapter 8: Crosslinguistic trends and variation Chapter 8 is split into four sections that address crosslinguistic trends with regard to the following four properties: article inflection, article paradigms, the distribution of different article types and the distribution of different article systems.

A look at the crosslinguistic distribution of article inflection in the section 8.1 reveals that the inflectional properties of articles mostly depend on the inflectional morphology on the noun that the article occurs with and the inflectional properties of the article’s source element. Section 8.2 deals with article paradigms. I discuss the paradigm of the definite article in Mokpe which is an example for a particularly complex paradigm. It illustrates how different noun classes can trigger the article to surface as a segmental or as a tonal article marker, but that they have to be two exponents of a single definite article because of their use in the same referential contexts.

Turning to the crosslinguistic distribution of article types in section 8.3 of this chapter reveals three main areal trends: a high number of definite articles in Africa, a relatively high number of indefinite articles in Eurasia and South America, and a high number of anaphoric articles in Australia. In section 8.3.3, I discuss one rarely attested and one unattested article type: non-anaphoric definite articles and generic articles. I propose two scenarios that can lead to the development of non-anaphoric definite articles and argue that most of the reported examples are not non-anaphoric articles in the strict sense. Generic articles, on the other hand, are not attested. I show a number of examples of other article types that can be used to mark generic referents.

In the last part of this chapter, section 8.4, I propose a typology of article systems according to the number of referent types that are encoded. The types of article systems represent the possible combinations of articles that encode one, two, or three referent types. I show and discuss the frequency distributions of the attested systems and crosslinguistic gaps, relating them to two main general principles.
1.3 Data and methodology

The major part of this study is based on a sample of 113 languages with 148 articles in total. Since I only included languages that have articles, the sample is a convenience sample. Nevertheless, the sample contains between 15 and 22 languages for each of the six macroareas introduced in the World Atlas of Language Structure and Glottolog (Dryer & Haspelmath 2013, Hammarström et al. 2018). The macroareas are: Africa, Eurasia, Papunesian, Australia, North America, and South America. The languages in my sample are provided in Tables 1.3 to 1.8 at the end of this section, sorted by macroareas, language families, and listed together with the source(s) used in this study. Figure 1.1 shows the areal distribution of the languages in my sample. Each dot represents a language, the different colors refer to the macroareas.

![Figure 1.1: Areal coverage of the sample](image)

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 the WALS chapter on definite articles (Dryer 2013a) show this trend as well. Figure 1.2 below features the map that belongs to this chapter. Blue and red dots indicate definite articles, white diamonds correspond to languages with an indefinite article but no definite articles, and

---

1This division into six macroareas is not without its downsides, as discussed in Hammarström & Donohue (2014). Because my sample is a convenience sample constrained by the availability of articles in the world’s languages, and because areal effects are not the main interest of this study, this issue can be neglected.

2The map was created using the R package “lingtypology” (Moroz 2017) which imports the data from Glottolog (Hammarström et al. 2018).

3The map is available online at [http://wals.info/feature/37A](http://wals.info/feature/37A).
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 also based on a larger sample, Russia or rather North-East Eurasia in general seems to lack languages with articles.

![Areal distribution of definite and indefinite articles (Dryer 2013a)](image)

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

It also follows from this unequal distribution of articles in the world’s languages that the macroareas do not contain an equal number of languages; in the sample used in this study, South and North America as well as Papunesia are represented by a slightly lower number of languages than Eurasia and Africa, while Australia clearly has the lowest number of languages in the sample, which also follows from the trend against articles in that 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.\(^4\) This is the case for Niger-Congo, Indo-European, Austronesian, and Pama-Nyungan. Since the different languages from these families are generally diverse, especially with respect to their article related properties, it should not bias the sample with regard to the purposes of the present study. My sample also includes a number of languages from the Arawakan family. The latter is not as large 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 included 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, Mokpe, and Runyankore), Indo-Aryan (Domari, Palula, Rajbanshi), and Oceanic (Maori, Rapa Nui, Sunwadia, Siar Lak, Sye, and Tongan).

\(^4\)Large is meant here as containing a large number of languages (cf. [http://glottolog.org/glottolog/family](http://glottolog.org/glottolog/family)).
Oceanic, Bantu, and Indo-Aryan are very large genera, most of them covering a large geographical area and having different types of articles. The second point also holds for the two Kwa languages in the sample: they show different properties regarding their article systems, which is why both were included into the sample used in this study. 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 data that I collected. Table 1.2 gives an overview of the languages, the number, age, and sex of the consultants, as well as the locations and dates. In the following Chapters, I indicate primary data as “prim. data” in the examples. From the languages for which I have primary data, Mokpe, Limbum, and Akan are also part of the 113-languages sample. Most other languages, i.e. Macedonian, Kaqchikel, Q’anjobal, and Tz’utujil are not part of the sample but serve as illustrations in various parts of the study. The two remaining languages Ejagham and Tikuna do not have articles; I use data from these languages mainly in the first part of the study to illustrate certain other relevant properties.

The presentation of examples from languages other than English follows the Leipzig glossing rules. 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.

---

5I thank my colleagues Jude Nformi and Sampson Korsah for their help as informants and providing me with data from Limbum and Akan.

<table>
<thead>
<tr>
<th>language</th>
<th>location</th>
<th>time</th>
<th>informants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mokpe</td>
<td>Buea (Cameroon)</td>
<td>Feb 2017</td>
<td>female, 21, male</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ejagham</td>
<td>Buea (Cameroon)</td>
<td>Feb 2017</td>
<td>male, 28</td>
</tr>
<tr>
<td>Limbum</td>
<td>Buea (Cameroon)</td>
<td>Feb 2017</td>
<td>male, 29</td>
</tr>
<tr>
<td></td>
<td>Leipzig (Germany)</td>
<td>Feb 2018</td>
<td></td>
</tr>
<tr>
<td>Akan</td>
<td>Leipzig (Germany)</td>
<td>Jun 2017</td>
<td>male, 31</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Leipzig (Germany)</td>
<td>Apr-Jun 2017</td>
<td>female, 26</td>
</tr>
<tr>
<td>Tikuna</td>
<td>Bogotá (Colombia)</td>
<td>Oct 2017</td>
<td>male, 28, male, 25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaqchikel</td>
<td>Patzún (Guatemala)</td>
<td>Nov 2017</td>
<td>female, 36, female, 27, female, 65</td>
</tr>
<tr>
<td>Q’anjobal</td>
<td>Santa Eulália (Guatemala)</td>
<td>Nov 2017</td>
<td>male, 65, male, 41</td>
</tr>
<tr>
<td>Awakateko</td>
<td>Aguacatán (Guatemala)</td>
<td>Nov 2017</td>
<td>female, 65, female, 40</td>
</tr>
<tr>
<td>Tz’utujil</td>
<td>San António (Guatemala)</td>
<td>Dec 2017</td>
<td>female, 35, male, 38</td>
</tr>
</tbody>
</table>
Table 1.3: Languages from the macroarea “Africa”

<table>
<thead>
<tr>
<th>Language</th>
<th>Genus</th>
<th>Country</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Afro-Asiatic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Dime</td>
<td>South Omotic</td>
<td>Ethiopia</td>
<td>(Seyoum 2008)</td>
</tr>
<tr>
<td>2 Konso</td>
<td>East Cushitic</td>
<td>Ethiopia</td>
<td>(Orkaydo 2013)</td>
</tr>
<tr>
<td>3 Buwal</td>
<td>Chadic</td>
<td>Cameroon</td>
<td>(Viljoen 2013)</td>
</tr>
<tr>
<td>4 Sheko</td>
<td>Dizoid</td>
<td>Ethiopia</td>
<td>(Helltenthal 2010)</td>
</tr>
<tr>
<td><strong>Blue Nile Mao</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Bambassi</td>
<td>Blue Nile Mao</td>
<td>Ethiopia</td>
<td>(Ahland 2012)</td>
</tr>
<tr>
<td><strong>Creoles</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Angolar Creole</td>
<td>Creole</td>
<td>Sao Tome and Principe</td>
<td>(Lorenzino 1998)</td>
</tr>
<tr>
<td>7 Cape Verde Creole</td>
<td>Creole</td>
<td>Cape Verde</td>
<td>(Baptista 2002)</td>
</tr>
<tr>
<td><strong>Dogon</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Jamsay</td>
<td>Plains Dogon</td>
<td>Mali</td>
<td>(Heath 2008)</td>
</tr>
<tr>
<td><strong>Eastern Sudanic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Gaahmg</td>
<td>Eastern Jebel</td>
<td>Sudan</td>
<td>(Stirtz 2011)</td>
</tr>
<tr>
<td>10 Lango</td>
<td>Nilotic</td>
<td>Uganda</td>
<td>(Noonan 1992)</td>
</tr>
<tr>
<td><strong>Mande</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Beng</td>
<td>Eastern Mande</td>
<td>Côte d’Ivoir</td>
<td>(Paperno 2014)</td>
</tr>
<tr>
<td><strong>Niger-Congo</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Akan</td>
<td>Kwa (Nyo)</td>
<td>Ghana</td>
<td>(Amfo 2010), (Arkoh 2011), (Arkoh &amp; Matthewson 2013), Sampson Korsah, pc</td>
</tr>
<tr>
<td>13 Balanta Ganja</td>
<td>North-Central Atlantic</td>
<td>Guinea-Bissau, Senegal</td>
<td>(Fudeman 1999)</td>
</tr>
<tr>
<td>14 Bullom So</td>
<td>Mel</td>
<td>Guinea, Sierra Leone</td>
<td>(Childs 2011)</td>
</tr>
<tr>
<td>15 Limbum</td>
<td>Bantu (Narrow Grassfields)</td>
<td>Cameroon</td>
<td>(Fransen 1995), Jude Nformi, pc</td>
</tr>
<tr>
<td>16 Logba</td>
<td>Kwa (Na-Togo)</td>
<td>Ghana</td>
<td>(Dorvol 2008)</td>
</tr>
<tr>
<td>17 Mokpe</td>
<td>Bantu (Sawabantu)</td>
<td>Cameroon</td>
<td>primary data, (Atindogbe 2013)</td>
</tr>
<tr>
<td>18 Oko</td>
<td>Benue-Congo</td>
<td>Nigeria</td>
<td>(Atoyebi 2010)</td>
</tr>
<tr>
<td>19 Runyankore</td>
<td>Bantu</td>
<td>Uganda</td>
<td>(Asimwe 2014)</td>
</tr>
<tr>
<td>20 Supyire</td>
<td>Gur</td>
<td>Mali</td>
<td>(Carlson 1994)</td>
</tr>
<tr>
<td><strong>Sandawe</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 Sandawe</td>
<td>Sandawe</td>
<td>Tanzania</td>
<td>(Steeman 2011)</td>
</tr>
<tr>
<td><strong>Songhay</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 Koyra Chiini</td>
<td>Northwest Songhay</td>
<td>Mali</td>
<td>(Heath 1999)</td>
</tr>
<tr>
<td>Language</td>
<td>Genus</td>
<td>Country</td>
<td>Reference</td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>--------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td><strong>Afro-Asiatic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Barwar</td>
<td>Semitic</td>
<td>Iraq</td>
<td>(Khan 2008)</td>
</tr>
<tr>
<td>Neo-Aramaic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mongolic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Bonan</td>
<td>Mogholi</td>
<td>China</td>
<td>(Fried 2010)</td>
</tr>
<tr>
<td><strong>Austro-Asiatic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Kharia</td>
<td>Munda</td>
<td>India</td>
<td>(Peterson 2011)</td>
</tr>
<tr>
<td><strong>Austronesian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Sri Lanka Malay</td>
<td>Malayo-Sumbawan</td>
<td>Sri Lanka</td>
<td>(Nordhoff 2009)</td>
</tr>
<tr>
<td><strong>Dravidian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Tamil</td>
<td>South Dravidian</td>
<td>India</td>
<td>(Lehmann 1993)</td>
</tr>
<tr>
<td><strong>Indo-European</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Albanian</td>
<td>Albanian</td>
<td>Albania</td>
<td>(Newmark et al. 1982)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Buchholz &amp; Fiedler 1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Zymberi 2004)</td>
</tr>
<tr>
<td>7 Armenian</td>
<td>Armenic</td>
<td>Armenia</td>
<td>(Düm-Tragut 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Andonian 2005)</td>
</tr>
<tr>
<td>8 Domari</td>
<td>Indo-Aryan</td>
<td>Jordan, Iran, Isrl</td>
<td>(Matras 2012)</td>
</tr>
<tr>
<td>(Central Zone)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 German</td>
<td>Germanic</td>
<td>Germany</td>
<td>(Hentschel &amp; Weydt 2003)</td>
</tr>
<tr>
<td>10 Irish</td>
<td>Celtic</td>
<td>Ireland</td>
<td>(Stenson 2008a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Stenson 2008b)</td>
</tr>
<tr>
<td>11 Palula</td>
<td>Indo-Aryan</td>
<td>Pakistan</td>
<td>(Liljegren 2016)</td>
</tr>
<tr>
<td>(Dardic)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Rajbanshi</td>
<td>Indo-Aryan</td>
<td>Bangladesh, India</td>
<td>(Wilde 2008)</td>
</tr>
<tr>
<td>(Eastern Zone)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Isolate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Basque</td>
<td>Isolate</td>
<td>Spain, France</td>
<td>(de Rijk 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Hualde &amp; Ortiz de Urbina 2003)</td>
</tr>
<tr>
<td><strong>Nakh-Dagestanian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Aghul</td>
<td>Lezgic</td>
<td>Russia</td>
<td>(Ganenkov et al. 2009)</td>
</tr>
<tr>
<td><strong>Sino-Tibetan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Ao (Mongsen)</td>
<td>Kuki-Chin-Naga</td>
<td>India</td>
<td>(Coupe 2007)</td>
</tr>
<tr>
<td>16 Bantawa</td>
<td>Kiranti</td>
<td>Nepal</td>
<td>(Doornenbal 2009)</td>
</tr>
<tr>
<td>17 Hakhun Tangsa</td>
<td>Brahmaputran</td>
<td>India</td>
<td>(Boro 2017)</td>
</tr>
<tr>
<td>18 Lepcha</td>
<td>Tibeto-Burman</td>
<td>Bhutan, India, Nepal</td>
<td>(Plaisir 2006)</td>
</tr>
<tr>
<td>19 Kurtöp</td>
<td>Bodic</td>
<td>Bhutan, China</td>
<td>(Hyslop 2011)</td>
</tr>
<tr>
<td><strong>Turkic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 Turkish</td>
<td>Oghuz</td>
<td>Turkey</td>
<td>(Göksel &amp; Kerslake 2005)</td>
</tr>
<tr>
<td><strong>Uralic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 Hungarian</td>
<td>Ugric</td>
<td>Hungary</td>
<td>(Kenesei et al. 1998)</td>
</tr>
</tbody>
</table>
**Table 1.5: Languages from the macroarea “Papunesia”**

<table>
<thead>
<tr>
<th>Language</th>
<th>Genus</th>
<th>Country</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Angan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Angan Menya</td>
<td>Angan</td>
<td>Papua New Guinea</td>
<td>(Whitehead 2006)</td>
</tr>
<tr>
<td><strong>Austro-Asiatic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Papuan Malay</td>
<td>Malayo-Sumbawan</td>
<td>Papua New Guinea</td>
<td>(Kluge 2017)</td>
</tr>
<tr>
<td><strong>Austronesian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Agta Northern</td>
<td>Luzon</td>
<td>Philippines</td>
<td>(Robinson 2008)</td>
</tr>
<tr>
<td>4 Bajau Basap</td>
<td>Greater Barito</td>
<td>Malaysia</td>
<td>(Miller 2007)</td>
</tr>
<tr>
<td>5 Biak South</td>
<td>Halmahera-West</td>
<td>Indonesia</td>
<td>(van den Heuvel 2006)</td>
</tr>
<tr>
<td>6 Maori Oceanic</td>
<td>Eastern Polynesian</td>
<td>New Zealand</td>
<td>(Harlow 2012)</td>
</tr>
<tr>
<td>12 Urama</td>
<td>Northeast Kiwai</td>
<td>Papua New Guinea</td>
<td>(J. Brown et al. 2016)</td>
</tr>
<tr>
<td><strong>Morehead-Wasur</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Komnzo</td>
<td>Tonda</td>
<td>Papua New Guinea</td>
<td>(Döhler 2018)</td>
</tr>
<tr>
<td><strong>Nuclear Trans New Guinea</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Oksapmin Ok-Oksapmin</td>
<td>Papua New Guinea</td>
<td>(Loughnane 2009)</td>
<td></td>
</tr>
<tr>
<td><strong>Solomons East Papuan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Bilua Bilua</td>
<td></td>
<td>Solomon Islands</td>
<td>(Obata 2003)</td>
</tr>
<tr>
<td>16 Lavukaleve</td>
<td>Lavukaleve</td>
<td>Solomon Islands</td>
<td>(Terrill 2003)</td>
</tr>
<tr>
<td><strong>Timor-Alor-Pantar</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Abui Alor</td>
<td></td>
<td>Indonesia</td>
<td>(Kratochvil 2007)</td>
</tr>
<tr>
<td>18 Bunaq East Timor Bunaq</td>
<td>Indonesia, Timor Leste</td>
<td>(Schapper 2009)</td>
<td></td>
</tr>
<tr>
<td>19 Teiwa Teiwa-Sar</td>
<td></td>
<td>Indonesia</td>
<td>(Klamer 2010)</td>
</tr>
<tr>
<td>Language</td>
<td>Genus</td>
<td>Country</td>
<td>Reference</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>---------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Bunuban</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Gooniyandi</td>
<td>Bunuban</td>
<td>Australia</td>
<td>(McGregor 1990)</td>
</tr>
<tr>
<td>Gunwinyguan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Bininj Kun-Wok</td>
<td>Marne</td>
<td>Australia</td>
<td>(Evans 2003)</td>
</tr>
<tr>
<td>3 Wubuy</td>
<td>Gunwinyguan Bak</td>
<td>Australia</td>
<td>(Heath 1984)</td>
</tr>
<tr>
<td>Mangarrayi-Maran</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Mangarrayi</td>
<td>Mangarrayi</td>
<td>Australia</td>
<td>(Merlan 1989)</td>
</tr>
<tr>
<td>Maningrida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Guragone</td>
<td>Bureran</td>
<td>Australia</td>
<td>(R. Green 1995)</td>
</tr>
<tr>
<td>Mirndi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Jingulu</td>
<td>Djingili</td>
<td>Australia</td>
<td>(Pensalfini 1997)</td>
</tr>
<tr>
<td>Nyulnyulan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Bardi</td>
<td>Western Nyulnyulan</td>
<td>Australia</td>
<td>(Bowern 2004, 2008)</td>
</tr>
<tr>
<td>8 Warrwa</td>
<td>Eastern Nyulnyulan</td>
<td>Australia</td>
<td>(McGregor 1994)</td>
</tr>
<tr>
<td>Pama-Nyungan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Arrernte (Mparntwe)</td>
<td>Arandic-Thura-Yura</td>
<td>Australia</td>
<td>(Wilkins 1989)</td>
</tr>
<tr>
<td>10 Diyari</td>
<td>Karnic</td>
<td>Australia</td>
<td>(Austin 2011)</td>
</tr>
<tr>
<td>11 Martuthunira</td>
<td>South-West Pama-Nyungan</td>
<td>Australia</td>
<td>(Dench 1994)</td>
</tr>
<tr>
<td>12 Yankuntjatjara</td>
<td>Desert Nyungic</td>
<td>Australia</td>
<td>(Goddard 1985)</td>
</tr>
<tr>
<td>13 Yolnu</td>
<td>Yimidhirr-Yalanji-Yidinic</td>
<td>Australia</td>
<td>(Wilkinson 1991)</td>
</tr>
<tr>
<td>14 Yuwaalaraay</td>
<td>Southeastern Pama-Nyungan</td>
<td>Australia</td>
<td>(Giacon 2014, Williams 1980)</td>
</tr>
<tr>
<td>Yangmanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Wardaman</td>
<td>Yangmanic</td>
<td>Australia</td>
<td>(Merlan 1994)</td>
</tr>
<tr>
<td>Language</td>
<td>Genus</td>
<td>Country</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Algic</td>
<td>Yurok</td>
<td>USA</td>
<td>(Garrett 2010, Robins 1958)</td>
</tr>
<tr>
<td>Chumash</td>
<td>Chumash</td>
<td>USA</td>
<td>(Henry 2012)</td>
</tr>
<tr>
<td>Haidan</td>
<td>Haida</td>
<td>Canada, USA</td>
<td>(Enrico 2003, Lachler 2004)</td>
</tr>
<tr>
<td>Hokan</td>
<td>Chimariko</td>
<td>USA</td>
<td>(Jany 2009)</td>
</tr>
<tr>
<td>Mayan</td>
<td>Ch’ol</td>
<td>Mexico</td>
<td>(Vázquez Alvarez 2011)</td>
</tr>
<tr>
<td>Misulmalpan</td>
<td>Matagalpan</td>
<td>Nicaragua</td>
<td>(Barlow 2018, T. Green 1999)</td>
</tr>
<tr>
<td>Mixe-Zoque</td>
<td>Zoque</td>
<td>Mexico</td>
<td>(Johnson 2000)</td>
</tr>
<tr>
<td>Muskogean</td>
<td>Alabaman-Koasati</td>
<td>USA</td>
<td>(Kimball 1985)</td>
</tr>
<tr>
<td>Otomanguean</td>
<td>Zapotecan</td>
<td>Mexico</td>
<td>(McIntosh 2011)</td>
</tr>
<tr>
<td></td>
<td>Chocho-Popolocan</td>
<td>Mexico</td>
<td>(Campbell &amp; Luna Villanueva 2011)</td>
</tr>
<tr>
<td>Penutian</td>
<td>Sahaptian</td>
<td>USA</td>
<td>(Jansen 2010)</td>
</tr>
<tr>
<td>Salishan</td>
<td>Central Salish</td>
<td>Canada</td>
<td>(Suttles 2004)</td>
</tr>
<tr>
<td>Siouan</td>
<td>Lakota</td>
<td>USA</td>
<td>(Van Valin 1977)</td>
</tr>
<tr>
<td></td>
<td>Mississippi Valley</td>
<td>USA</td>
<td>(Williamson 1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Rood &amp; Taylor 1996)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Ingham 2003)</td>
</tr>
<tr>
<td>Totonacan</td>
<td>Tepehua</td>
<td>Mexico</td>
<td>(Kung 2007)</td>
</tr>
<tr>
<td>Uto-Aztecan</td>
<td>California Uto-Aztecan</td>
<td>USA</td>
<td>(Hill 2006)</td>
</tr>
<tr>
<td></td>
<td>Aztec</td>
<td>El Salvador</td>
<td>(Campbell 1985)</td>
</tr>
<tr>
<td></td>
<td>Tarahumaran</td>
<td>Mexico</td>
<td>(Cohen 1998)</td>
</tr>
<tr>
<td></td>
<td>Numic</td>
<td>USA</td>
<td>(Givón 2011)</td>
</tr>
<tr>
<td>Wakashan</td>
<td>Nootkan</td>
<td>Canada</td>
<td>(Davidson 2002)</td>
</tr>
</tbody>
</table>
Table 1.8: Languages from the macroarea “South America”

<table>
<thead>
<tr>
<th>language</th>
<th>genus</th>
<th>country</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Araucanian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Mapudungun</td>
<td>Araucanian</td>
<td>Chile</td>
<td>(Smeets 2008)</td>
</tr>
<tr>
<td><strong>Arawakan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Arawak</td>
<td>Island Carib-Garifuna</td>
<td>Suriname</td>
<td>(Pet 2011)</td>
</tr>
<tr>
<td>3 Asheninka Peréné</td>
<td>Ashe-Asha Norte</td>
<td>Peru</td>
<td>(Mihas 2010)</td>
</tr>
<tr>
<td>4 Bare</td>
<td>Inland Northern Maipuran</td>
<td>Brazil, Venezuela</td>
<td>(Aikhenvald 1995)</td>
</tr>
<tr>
<td>5 Baure</td>
<td>Bolivia-Parana</td>
<td>Bolivia</td>
<td>(Danielsen 2007)</td>
</tr>
<tr>
<td>6 Parecis</td>
<td>Paresi-Saraveka-Saluma</td>
<td>Brazil</td>
<td>(Barros Brandão 2014)</td>
</tr>
<tr>
<td>7 Wayuu</td>
<td>Guajiro-Paraujan</td>
<td>Colombia</td>
<td>(Álvarez 2017)</td>
</tr>
<tr>
<td><strong>Cariban</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Carib</td>
<td>Cariban</td>
<td>Suriname</td>
<td>(Courtz 2008)</td>
</tr>
<tr>
<td><strong>Chicham</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Aguaruna</td>
<td>Chicham</td>
<td>Peru</td>
<td>(Overall 2007)</td>
</tr>
<tr>
<td><strong>Tupian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Akuntsú</td>
<td>Tuparic</td>
<td>Brazil</td>
<td>(Aragon 2014)</td>
</tr>
<tr>
<td>11 Nhengatu</td>
<td>Tupi-Guarani</td>
<td>Brazil, Colombia</td>
<td>(da Cruz 2011)</td>
</tr>
<tr>
<td><strong>Macro-Ge</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Apinayé</td>
<td>Core Northern Ge</td>
<td>Brazil</td>
<td>(Cunha de Oliveira 2005)</td>
</tr>
<tr>
<td><strong>Nambikwaran</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Mamaindé</td>
<td>Nambikwara Complex</td>
<td>Brazil</td>
<td>(Eberhard 2009)</td>
</tr>
<tr>
<td>14 Sabanè</td>
<td>Sabanè</td>
<td>Brazil</td>
<td>(Antunes de Araujo 2004)</td>
</tr>
<tr>
<td><strong>Pano-Tacanan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Kashibo-Kakataibo</td>
<td>Panoan</td>
<td>Peru</td>
<td>(Zariquiey Biondi 2011)</td>
</tr>
<tr>
<td><strong>Tucanoan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Cubeo</td>
<td>Cubeo-Desano</td>
<td>Colombia</td>
<td>(Chacon 2012)</td>
</tr>
<tr>
<td><strong>Zamucoan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Ayoreo</td>
<td>Zamuco-Ayoreo</td>
<td>Bolivia, Paraguay</td>
<td>(Bertinetto 2009, Ciucci 2016)</td>
</tr>
</tbody>
</table>
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 (1), illustrating the definite and indefinite articles in English in opposition to each other.

(1)   a. The woman got off the train.
     b. The woman got off a train.

Glossing over the details for now, we can say that the definite article in (1a) marks the referent of train as identifiable, known, given, or accessible, and also as prominent in the scenario described, while the indefinite article in (1b) 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 proximity to the articles found in English.

In the following paragraphs, I will present a number of examples that illustrate this problem. Examples (2) and (3) 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 (3) as an indefinite article.

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

(3) ta qunian mai le [(yi) zhuang fangzi]
    3sg last.year buy PFV one CL house
    ‘She/he bought a house last year.’                           Chinese (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. Chen 2003). It is even attested in predicative contexts which are no longer referential:

(4) ta shi [(yi) ge mainairen]
    3sg be one CL businessman
    ‘She is a businessman.’                                      Chinese (Chen 2003: 1171)

Another example of markers whose classification is problematic are the possessive markers in a number of Uralic languages, which have functions very similar to definite articles. This is what we find in, for instance, Udmurt: although still recoverable as the third person singular possessive marker, -ez, was argued to indicate definiteness even in contexts with no possessor being semantically present or recoverable (e.g. Gerland 2014):\(^1\)

(5) so-len eš-ez
    he-GEN friend-POSS:3SG
    ‘his friend’                                                    Udmurt (Edygarova 2009: 101)

---

\(^1\)I do not discuss here whether or not this possessive marker in Udmurt should be treated as an article. In general, there is no consensus to what extent possessives in Uralic classify as definite articles or express definiteness based on possession. Fraurud (2001), Nikolaeva (2003) argued against an account as a definiteness marker because they take the marking to be optional, while Csúcs (2003), É. Kiss & Tánecz (2017), Gerland (2014) base their analyses on possessives as indicators of definiteness and analyse definiteness marking as generally required but allowing for exceptions.
In (5), we see -ez in its original function, marking the possessor on the possessee es ‘friend’. In (6) and (7), on the other hand, the contexts do not include a potential possessor, and the same marker -ez is used to indicate that the referents guy and grass are marked as identifiable and hence as definite. Again, this shows that one needs language-independent criteria to classify a marker as an article.

Another entirely different phenomenon that touches upon the definition of articles as a cross-linguistic 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 (8) from Nivaclé (Matacoan, Paraguay) shows how different past markers on the noun mark the referent as anaphoric and thus as definite. In (8a), -naxi indicates yesterday’s past and -mati is used in (8b) to mark today’s past.

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.

The last three examples from Chinese, Udmurt, and Nivaclé showed that it is not sufficient to take the properties of the “European” article as a comparative concept of articles. 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. Example (9) and (10) show German and English, which do not use a definite article with a generic referent in the plural, while (11) and (12) show that Spanish and Hungarian require it:

(9) I like (*the) cats.
    English (prim. data)

(10) Ich mag ([*die] Katzen].
    1sg like.1sg ART:DEF cats
    'I like cats.'
    German (prim. data)

(11) Me gustan [*(los) gatos].
    1sg.dat please.3pl ART:DEF cats
    'I like cats.'
    Spanish (prim. data)

(12) Szeretem [*a) macskákat].
    like.1sg ART:DEF cats
    'I like cats.'
    Hungarian (prim. data)

There are two main problems in the perception of articles. On the one hand, we see that even in European languages, the distribution of articles varies, and a distributional criterion for articles must take this into account. On the other hand, the fact that definite articles occur with generic referents also shows that the use of articles is not necessarily restricted to their referential values.

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 Introduction

A suitable approach to define articles for crosslinguistic comparison is the one of comparative concepts, proposed in Haspelmath (2010).² Comparative concepts “are not part of particular language systems and are not needed by descriptive linguists or by speakers. They are not psychologically real, and they cannot be right or wrong. They can only be more or less well suited to the task of permitting crosslinguistic comparison […] Comparative concepts are universally applicable, and they are defined on the basis of other universally applicable concepts: universal conceptual-semantic

---

²The 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).
concepts, general formal concepts, and other comparative concepts” (Haselma 2010: 665). In the present study, I will adopt this approach to define a crosslinguistic category of articles. That we are still in need of a detailed discussion can be shown by previous definitions of articles. For instance, even though the definition of articles in Dryer (2014: 234) is similar to what I propose in this study, it is formulated in rather vague way which makes it difficult to apply consistently to different languages.

[...]

Another definition of articles is given in Lyons (1999). In his approach to articles, he emphasizes that the function of articles should not be restricted to the marking of (in)definiteness.

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. However, the latter are formulated in a vague way that does not make it explicit how they can be applied in practice.

In contrast to these two definitions that do not distinguish between different types of articles, Schroeder (2006) provides two separate definitions for definite and indefinite articles, respectively, referring to their formal and functional properties. In these two definitions, the functional characteristics of the articles are rather vague way as well.

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)

This shows that there is still need for an explicit and applicable definition of articles. The goal of this section is to define articles as a crosslinguistic category following the approach of comparative concepts. The definition of articles as a crosslinguistic category is 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. I propose three main criteria to define articles. 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 nouns. 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 (13).

(13) **Articles as a crosslinguistic category**

    Articles are markers in the nominal domain whose main function it is to encode the referential function of the noun 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 referent types

Articles encode the referential functions of the noun that they occur with. I call these functions referent types. In this section, I mention a number of referent types that different types of articles encode (for a detailed definition of referent types and article types, see sections 3.2 and 3.3). In addition, I discuss a number of markers which also indicate referentiality but which are not classified as articles because their primary function is a different one.

We can distinguish four main referent types: definite, specific, nonspecific, and generic referents (cf. section 3.2). 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 requires a link between the referent to an object that is present in the discourse situation:

(14) **Do you see the house** over there?  

Anaphoric referents, on the other hand, are fully identifiable by the speaker and the hearer due to previous mentioning in the discourse. In (15), the referent of shoes is used anaphorically in B’s answer and can therefore be marked by the definite article the:

(15) **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.  

Similar to anaphoric contexts are bridging contexts which have also been labelled associative anaphoras because the two referents do not share full identity but are linked to each other in an unambiguous way. A typical example is the relation between a book and its author:
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 by A to be marked with the definite article.

A: What happened to the dog (we used to have)?

B: We had to give it away because your sister was allergic.

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 contextually unique in the context of a flat. In (19), on the other hand, the referent of moon is absolutely unique in our default perception of the world, which also makes it definite. As can be seen in (18) and (19) below, both types of unique referents require the use of the definite article in English.

18) Where is the kitchen in your flat?

19) The moon is beautiful tonight.

Another type of definite referents that has to be distinguished is the so-called establishing referent. Establishing referents go back to Hawkins (1978) and correspond to referents that are already marked as identifiable to the hearer, even though the referent is only made identifiable in the course of the immediately preceding utterance. An example is given in (20); again, the definite article is required to indicate the identifiability or definiteness of the referent of rumour.

20) London has been buzzing with the rumour that the Prime Minister is going to resign.

(Hawkins 1978: 102)

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 is linked to a particular referent of its kind. Nonspecificity, in contrast to definiteness and specificity, does not involve a particular referent that is linked to the noun. 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

---

3We can further distinguish two types of specific referents: referents that are only nonidentifiable for the hearer but not for the speaker, and referents that are nonidentifiable for either discourse participants. Since this distinction does not seem to be relevant for articles, I do not consider it in this study.
is linked. In other words, the discourse situation does not provide enough information to decide which of all the referents of a kind is linked to the noun. 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 (21), taken from Givón (1984: 441) and going back to Quine (1953):

(21) 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 (21a) and (21b). In (21a), 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 the speaker refers to. In (21b), 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.

Indefinite articles that mark both specific and nonspecific referents are also found in other parts of the world. The following examples show the indefinite article in Awakateko (Mayan, Guatemala). In contrast to the specific article jun in Q’anjobal, the article jun in Awakateko is an indefinite article because it marks both specific and nonspecific referents. In (22), the article encodes a specific referent, since only the speaker and not the hearer can unambiguously identify the church, but it is a particular church. As (23) shows, the same article jun is used for nonspecific referents which no longer correspond to any particular referent from their kind.

(22) at [jun  tiox]  stzi’  a’
    exist ART:INDEF church river.bank water
    ‘There is a church at the river banks.’
    Awakateko (prim. data)

(23) nawaj  [jun  moy]
    want.1sg ART:INDEF car
    ‘I want to have a (any) car.’
    Awakateko (prim. data)

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.\(^4\) The inclusive-specific article encodes definite as well as specific indefinite referents, but not nonspecific referents. As (24a) shows for Bemba, the referent of 'book' is necessarily interpreted as definite or specific when it is marked by \(i\)- as in \(icítabo\). When the referent is expressed by the noun without this marker, e.g. \(cítabo\) in (24b), it must be interpreted as nonspecific.\(^5\)

(24) 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 (1) and (2) below show for nouns of other classes (i.e. genders) 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.

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

(26) 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. Most probably for this reason, 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 illustrated in (27):

\(^4\)For a discussion of the inclusive-specific article, see section 6.1.

\(^5\)A number of Bantu languages feature this so-called “pre-prefix” or “augment” preceding the noun class prefix. The 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).
Since these two markers occur with different types of nouns, it seems plausible to call them articles. However, their function is different: *e* occurs only with proper nouns referring to persons as in (27a) and *na* is used only with common nouns as in (27b). This suggests that their function is not linked to referentiality, which is why I do not treat these markers as articles.

Other elements that occur with nouns but do not have primarily referential functions are case markers. An example is the accusative-partitive alternation in Finnish. In Finnish, case markers do not directly encode the referent type of the noun that they occur with, but they can influence the referential interpretation of nouns in certain syntactic positions. The accusative-partitive case alternation can influence the referential interpretation of the object of a transitive clause. An object marked by the accusative case as in (28a) is interpreted as having a definite referent, while the referent of an object in the partitive as in (28b) is interpreted as specific. Needless to say, such markers are not considered as articles because marking referentiality is not their primary function.

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. Example (29) and (30) illustrate this:
In (30) we can see that the lexical meaning of “deceased” is extended to mark a former state of the referent, ‘the English language’. The marker ē:to is treated as an article in the grammar for language-specific reasons.6 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 ē: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. I use “nominal domain” and “noun phrase (NP)” interchangeably, ignoring framework-dependent concepts, e.g. noun phrase vs. determiner phrase.

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-‘the cat’. If the article is affixal, 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 a number of nouns that used to be compound forms. An example is waki-‘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

6The 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.
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 (31) to (34) give an overview of the possible orders of the noun, the article, and the adjective:

- **Preposed and N-anchored**
  
  (31) büyük *bir* oda
  
  large  ART:INDEF room
  
  ‘a large room’
  
  Turkish (Göksel & Kerslake 2005: 145)

- **Preposed and NP-anchored**
  
  (32) juu maqaniyaa lapanak
  
  ART:DEF old  people
  
  ‘the old people’
  
  Tepehua (Kung 2007: 388)

- **Postposed and N-anchored**
  
  (33) kil ticé tilòlìyén
  
  houses ART:DEF low
  
  ‘the low houses’
  
  Bullom So (Childs 2011: 71)

- **Postposed and NP-anchored**
  
  (34) kavayi sāgharumu ’uway
  
  horse white ART:DEF
  
  ‘the white horse’
  
  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 classified as affixal 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). Although described as free variation in Nordhoff (2009: 319-320), the position of the article may be conditioned by the presence or absence of other elements that occur in the NP, as e.g. the additive marker in (35b).
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. sections 4.1.2 and 8.2.2 on the definite article in Mokpe). This is shown in (36) for a number of nouns in isolation, and for the noun likândô ‘roof’ in a bridging context in (20).

(36) Mokpe (prim. data)

a. móléli ‘food’ vs. móléli ‘the food’
   b. likálà ‘(a) bridge’ vs. likálà ‘the bridge’
   c. bètàngùlè ‘(some) lizards’ vs. bètàngùlè ‘the lizards’

In the following paragraphs, I address 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 (38b), the agreement marker in the verbal complex triggers a definite interpretation of the referent of the object.

(38) Nyaturu (Hualde 1989: 182)

a. n-a-onaa mwalimu
   sm:1sg-pst1-see cl1.teacher
   ‘I saw a teacher.’
b. n-a-mʊ-onaa mwalimu
   SM:1SG-PST1-OM: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. A well-known example is the German adjective agreement, which is sensitive to definite and indefinite contexts in the nominative:

(39) a. ein groß-er Hund
    ART:INDEF big-NOM.M.SG.INDEF dog
    ‘a big dog’

b. der groß-e Hund
    ART:DEF big-NOM.M.SG.DEF dog
    ‘the big dog’

Also in Khwarshi (Tzezic, Russia), we find a definiteness marker that only combines with adjectives and similar expressions. Accordingly, the marker so, as shown in (40) and (41), is not treated as a definite article.

(40) žik’we [miq’e-so baydan] m-ež-i
    man.OBL.ERG far.away-DEF field.CL3 CL3-plant-PST.W
    ‘The man planted the furthest field.’

(41) ø-øloł’o-so-ho y-øloł’o-so y-ez-un
    1-in.middle-DEF-APUD II-in.middle-DEF II-take-PST.UW
    ‘The middle (brother) married the (other) middle (sister).’

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).7 Example (42) illustrates the use of this marker ta.

---

7Colloquial 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 (43a); it cannot mark the referent as definite if no adjective is present (43b). Therefore, I do not regard \( ta \) as a definite article.

(43)  a.  \( ta \) velika knjiga  
       \[ \text{DEF big book} \]  
       ‘the big book’

b.  *\( ta \) knjiga  
       \[ \text{DEF book} \]  
       intended meaning: ‘the book’

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:

(44)  a.  *the a boy

b.  *a the boy

This is also taken 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.

As will be shown in section 8.3.3.2, we find languages that allow the co-occurrence of a definite article and an anaphoric or a recognitional article. It seems plausible that this is possible because there is no semantic conflict, as both articles mark the referent as identifiable, with a functionally broader definite and a more restrictive anaphoric or recognitional article. Example (45) shows the co-occurrence of a definite article and an anaphoric article in Jamsay:
The combination of two articles can also be semantically motivated in a different way. (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 showed 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 (46) 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.

(46) \[nkɔrɔfo no \ bì] \ ka-a se wɔ̃ 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 (47) illustrates this:

(47) \[maame bì \ no] \ a ɔ-be-hwehwɛ-e 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 one 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. As will be shown in section 8.1.1, 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 (48) shows a noun with a definite

(45) tògù pó:ró tógó kó bèrɛ̂ː bɛ̀rɛ̂ː nû: [kó tòg kù́] úró
shed first shed.building 3SG.NHUM in 3PL enterPFV ART:ANA shed ART:DEF house
táná-ŋá mèyⁿ nîŋ yɔ́=kɔ̀
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)
Referent being marked by a possessive. The noun *libri* ‘the book’ contains the definite article. As example (49) shows, even if the indefinite article is used to mark the referent of the noun as non-identifiable, the definite article remains obligatory.

\[(48)\]  
\[\text{ky} \, \varepsilon\text{sh}\text{t} \, \text{lib}-i \, \text{im}\]
\[\text{this is} \, \text{book-ART:DEF POSS:1SG}\]
\[\text{‘This is my book.’}\]

Albanian (Zymberi 2004: 58)

\[(49)\]  
\[\text{nj} \, \varepsilon \, \text{mi}-\text{k-}^{(u)} \, \text{im}\]
\[\text{ART:INDEF friend-ART:DEF POSS:1SG}\]
\[\text{‘a friend of mine’}\]

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 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)

Another example showing this issue is found in Terrill (2003) concerning Lavukaleve (Lavukaleve, Solomon Islands). Lavukaleve has an indefiniteness marker which is not considered as an article in the grammar. The marker is estimated to be too infrequent and not obligatory in the sense that not all indefinite noun phrases occur with this marker. According to the author, the strongest argument against treating the marker *ro* as an article is that it can co-occur with the definite article in the language.

> […] *ro* is a marker of indefiniteness, but one would not want to call it an indefinite article, for a number of reasons. Firstly, it has the syntactic distribution of an adjective. Secondly, it does not have the frequency or obligatoriness which one might expect of an article; it is certainly far less frequent than the definite article. Thirdly, it is not obligatory in indefinite noun phrases, again unlike the definite article, which is obligatory in all definite noun phrases (unless there is a demonstrative to express the definiteness). Fourthly, and crucially, it can, under certain constrained circumstances, co-occur with the definite article. (Terrill 2003: 81)

There are at least three problematic assumptions in these quotes presented. Firstly, the properties of an article in a single language (e.g. English) cannot serve as basis for a crosslinguistic comparison. Secondly, articles are not “obligatory” in any language if one understands obligatory as
occurring in all contexts in which they are expected according to their referential function. Often
we find that the use of articles is restricted in many contexts in which they should mark the re-
ferent type of the noun (e.g. mass nouns, plural nouns, with certain adpositions, with certain case
markers). Thirdly, although definite and indefinite articles usually do not co-occur in European
languages, this cannot be taken as an argument as such for or against treating another marker as
an article. In fact, co-occurring articles are attested in a number of languages (see section 2.2.3).

As mentioned above, it is problematic to characterize the distribution of articles as obligatory if
obligatoriness cannot allow for exceptions. I will illustrate this with what I regard as an anaphoric
article in Hausa (Chadic, Nigeria). Hausa has a marker -n, shown in example (50), which started
out as an anaphoric article but which is more and more used in other definite contexts as well.
Example (51) shows a noun which is marked by -n and which has a definite referent because of
the following relative clause.

(50)  a. bákâ-n ‘the bow’
    b. manòmā-n ‘the farmer’

(51)    yaarò-n  dà  ya  tàfi
        boy-ART:ANA/DEF 3SG.M.PERF.REL leave
             ‘the boy who left’

Hausa (Newman 2000: 144)

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 an 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:

(52)  tûlû  yā  fashè
      pot  3SG.PERF break
      ‘The / a water pot broke.’

Hausa (Newman 2000: 143)

In example (52), tûlû ‘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 also influenced by factors other than refer-
entiality; 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.
Therefore, I propose to base the distributional criterion for articles on the notion of systemati-
city instead of obligatoriness, since the latter does not allow for exceptions and suggests that
the decision between treating a marker as an article or not can be made on the basis of single
examples. However, the presence and absence of articles can be conditioned by so many other
syntactic, semantic, morphological, and prosodic properties of the construction, so that individual
examples are not sufficient to classify a marker as an article.

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. syntac-
tic, 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 (53a) shows *Katze ‘cat’ marked as nonidentifiable by the indefinite article. Its plural
counterpart in (53b) is expressed as a bare plural noun *Katzen ‘cats’ because the indefinite article
is not available in this context.

   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 (prim. data)

The relevant argument to treat this marker in German as an indefinite article here is that even
though nouns can be interpreted as nonidentifiable 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 lan-
guages, 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 (54), but does not occur with plural nouns, shown in (55):

(54) ék-ká fáád-k’á-bààb ín-f yi-ərá k’ödäm-m-s yis there-LOC body-in-father wood-ART:DEF.M DIST.M-ACC.CONT qodama-ART:DEF-M DIST.M há-gišū-köb-t=á kēs-ā-m-ə s:3M=pull-take-ss=s:3M go.out.CAUS-PUT-IRR-STI ‘that qodama pulls out the wood which is there in the body’ Sheko (Hellenthal 2010: 141)

(55) akǹ sàm-əb fyāānū-s-kn ʒéenʃ kī-ət-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 -ʃ is only available for singular nouns, Hellenthal (2010: 136) analyzes 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 (56a) shows in opposition to (56b), does not occur with nouns whose first segment is a vowel:

(56)  

a. è paŋga ‘the sand’  
b. ∅ ikpā ‘(the) salt’  

Mokpe (prim. data)

Since this restriction of the definite article can be formulated as a simple rule (è → ∅ / _V), it does not pose a problem for treating the marker as an article.

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 (57a) shows this for the definite article in English which does not combine with proper nouns, while example (57b) illustrates the incompatibility of the English indefinite article with mass nouns.

(57)  
a. (*The) Steve baked (*the) Hannah a cake.  
b. He is drinking (*a) milk.  

English (prim. data)

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 gram-
mar (Orkaydo 2013: 97), while it is generally used in anaphoric and other definite contexts (Orkaydo 2013: 95-97). Example (58) illustrates the use of the definite article for referents that have already been mentioned in the discourse:

(58) kaasa-sit-n=in karmaa-siʔ ?iʔfj-ay
‘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 (59) shows an utterance that comes shortly after the one in (58); the referent of karmaa ‘lion’ is definite, but the definite article is not used in this instance.

(59) karmaa ka ʔarapea-asiʔ ʔaraa kaassuma=i kaassaaf-ay
lion and monkey-DEM on question=3 ask-PF.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 the marker 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). In examples (60) and (61), we saw the “standard” specific and nonspecific indefinite articles -m and -eem, respectively.

(60) dakáak-kaata-m húu-laa hii-k
bird-DIM-ART:SPEC come-ss reach-DECL
‘a bird came, it reached him’ Crow (Graczyk 2007: 228)

(61) axée baláyyikaashe diuup-eem alúutkaashe áppaa día-a-wa-ku-hee?
father bow two-ART:NSPEC arrow with make-CONT-1-give-AFF-Q
‘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 (62a) and (62b) illustrate. Example (62a) features a specific referent (bachee-lák ‘a man’), a referent that is not identifiable but unambiguously linked to the noun. Example (62b) shows the noun iisashpí-t-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.
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 for 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:

(63)  
(a) $hn$-$ou$-$orch$
    $in$:ART$^\indef$-become.secure
    ‘securely’
(b) $hn$-$ou$-$ou\\mbox{\textcircled{b}}$-$ou\\mbox{\textcircled{b}}$-$bol$
    $in$:ART$^\indef$-show.forth
    ‘openly’

Coptic, Sahidic dialect (Plumley 1948: 40)

Another syntactic use of articles can be illustrated with indefinite articles from Q’anjobal (Mayan, Guatemala). The language has an exclusive-specific article and classifiers, but no definite article. The exclusive-specific article as well as classifiers precede the noun. In order to use an adnominal demonstrative which follows the noun, Q’anjobal requires this prenominal position to be filled. The default marker filling that position is the classifier, as is illustrated in examples (64) and (65).

(64)  
$cham$ icham $tu$
   CL:MAN man DEM:DIST
   ‘that man’
   Q’anjobal (prim. data)

(65)  
$te$ $na$ $tu$
   CL:TREE house DEM:DIST
   ‘that house’
   Q’anjobal (prim. data)

\textsuperscript{8}The classifiers do not have (primary) referential functions.
If no classifier is available for the noun in question, the indefinite article is used instead. Examples (66)-(68) show the use of the exclusive-specific article together with the demonstrative. The prenominal position can also be filled by a combination of the article and the classifier, as in (69).

(66) jun ora tu’
\text{ART:EXSPEC hour DEM:DIST}
‘at that hour’
Q’anjobal (prim. data)

(67) jun b’eqan ti’
\text{ART:EXSPEC moment DEM:PROX}
‘at this moment’
Q’anjobal (prim. data)

(68) jun tz’uy ti’
\text{ART:EXSPEC bag DEM:PROX}
‘this bag’
Q’anjobal (prim. data)

(69) jun winaq unin ti’
\text{ART:EXSPEC CL:MAN child DEM:PROX}
‘this boy’
Q’anjobal (prim. data)

Without going into detail with regard to the distribution of the exclusive-specific article and classifiers, the important point is that the noun cannot 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. Tz’utujil (Mayan, Guatemala) has a marker ja(la), which can occur in different definite contexts. Example (70) shows $ja$ in an anaphoric context; in (71), we see it with a situationally unique referent.

(70) A: taq xinipoon p-jaay xintzu’ [jun i xoq] k’in ik’ee’ ak’alaa’
when arrive.pst.1sg loc-house see.pst.1sg \text{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.’
B: awtaquin ruwach [ja i xoq]?
know.pst.2sg eyes \text{DEF woman}
‘Did you know that woman?’
Tz’utujil (prim. data)

(71) b’anitzra k’ool wa [ja q’atb’al] tzii chpam jawa tinamet?
where exist q \text{DEF town.hall inside dem town}
‘Where is the town hall in this town?’
Tz’utujil (prim. data)
In addition, *ja* does not occur in indefinite contexts, as is illustrated for specific referents in (72) and for nonspecific referents in (73). In both of the contexts, Tz’utujil uses the indefinite article *jun* instead.

(72) k’oola [jun sepaneem] awyin tat
    exis.1sg ART:INDEF gift for.2sg you
    ‘I have a gift for you.’
    Tz’utujil (prim. data)

(73) k’ool [jun atz’ib’el]?
    exist ART:INDEF pen
    ‘Do you have a pen?’
    Tz’utujil (prim. 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 informants, 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 informants, 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 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 occurs in the nominal domain to mark definiteness. The fact that speakers have no clear preference in favor or against the use of *ja* with definite referents, I conclude at this point that *ja* in Tz’utujil is not distributed in a systematic way, at least not as a definite article.\(^9\)

### 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, giving a brief overview of the major article types that we find across the languages of the world. These article types are: definite, anaphoric, inclusive-specific, exclusive-specific, indefinite, and referential 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 showed that further requirements with regard to their form or position

---

\(^9\)In 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.
are not necessary. Section 2.2.4 was concerned with the distributional criterion of articles as a
crosslinguistic category. I argued that systematicity is a better criterion than obligatoriness, since
the latter does not allow for exceptions. However, I showed that we commonly find two types
of exceptions concerning the distribution of articles: articles can either be absent in contexts in
which their presence is expected based on their referential function or they can be present in refer-
entially unexpected contexts. Systematicity can capture this distribution, allowing for exceptions
that can be described as rules or strong correlations.
Chapter 3

Defining different types of articles

This chapter proposes a definition of each of the 8 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 important background concepts concerning referentiality. In section 3.2, I then define the relevant referential functions, called referent types which are then related to the definition of the 8 article types in section 3.3.

3.1 Referentiality and referent types

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. I call the values of referentiality, i.e. the functions that referring expressions have, referent types. In formal semantics, referentiality has usually been approached from the sentence level. The study of referentiality restricted to the sentence level is is traditionally located within semantics, also labelled 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 engage in the discussion of distinguishing these two levels of sentence and discourse referentiality and their location within semantics or pragmatics. In doing so, I follow Hawkins's agenda: “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

In this section, I introduce a schematic discourse situation in order to lay the ground for the discussion of referentiality and for the distinction between different referential functions (referent
types), which in turn is the basis for the distinction between different types of articles. The relevant notions are: the discourse situation, the participants (hearer and speaker), and the discourse referents. To illustrate how the discourse referents relate to the participants and the discourse itself, consider the following example in (1).

(1) Context: *S and H are chatting at the bus stop, waiting for the bus to come. A man approaches the bus stop. S is saying to H:*

   *S: That man is tall!*

The discourse situation in (1) can be characterized in terms of the context and the discourse participants, namely the speaker (S) and the hearer (H). The sentence uttered by the speaker contains the linguistic expression *that man*, which is linked to the discourse referent $R_{man}$. Figure 3.1 shows how such a discourse situation can be represented schematically.

For reasons of simplicity, I assume that a discourse situation involves two discourse participants, i.e. the speaker and the hearer. 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 *That man is tall!*

In order to properly distinguish between form and function, referents$^1$ are represented on a different level than the referring linguistic expressions. Referents are defined as (abstract) elements

---

$^1$I use the shorter expression “referents” in the meaning of “discourse referents”.


in what I call the speaker’s and the hearer’s mental spaces $M_S$ and $M_H$. The concept of “mental space” goes back to Fauconnier (1994, 2007) and represents 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 (1), the referents $R_{man}$, $R_{bus, stop}$, and $R_{bus}$ are part of both the speaker’s mental space $M_S$ and the hearer’s mental space $M_H$ because they are 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 (1), the referent $R_{man}$ is linked to the expression *that man*.

As is schematically illustrated in Figure 3.1, the intersection of their two mental spaces $M_S$ and $M_H$ contains the discourse universe $D$ which contains all the referents from the discourse that both the speaker and the hearer can identify. The relation between the two mental spaces and the discourse universe thus is the following:

\[(2) \quad D \subseteq (M_S \cap M_H)\]

The discourse universe is a subset of the intersection of the speaker’s and hearer’s mental spaces $M_S$ and $M_H$ because both discourse participants might share knowledge about other referents based on previous shared experience or general and world knowledge. Hence, the discourse universe may not contain all the referents that are mutually identifiable by the speaker and the hearer. This will be shown to be important for the definition of two referent types that are relevant to articles (cf. section 3.2).

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).

The separation of the linguistic expression and the referent is crucial for the discussion of articles as linguistic expressions that encode referential functions.\(^2\) If not distinguished properly, the crosslinguistic comparison of articles is problematic. For instance, in their influential paper on cognitive statuses of discourse referents, Gundel et al. (1993: 275) directly map referential

\(^2\)Krifka & Musan (2012a: 6) made this point on a related note: “For some reason, this confusion of expression and meaning occurs particularly often with information structural notions. With notions like subject, predicate or direct object it does not arise; no one would claim that John the person is the grammatical subject [...], it is John the noun phrase. The imprecision of information structure terms can be endured if one is aware of it.”
functions onto linguistic expressions in English. Therefore, the referential functions rely on linguistic expressions in a single language and cannot be applied directly to other languages. For this reason, I define referential functions form-independently as referent types in section 3.2.

Modelling discourse referents as abstract objects in an abstract space (here: the discourse universe) is not a new approach. Going back to Frege (1892), propositions as events about entities (referents) are evaluated as true or false against the real world in formal semantics and logic. 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”. 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.\footnote{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. To illustrate this, he provides the following examples:}

\begin{align*}
(3) & \quad \text{a. The present } \textbf{king of France} \text{ is not bald.} \\
& \quad \text{b. The present } \textbf{queen of England} \text{ is not bald.} \\
& \quad \text{English (Givón 1984: 438)}
\end{align*}

\begin{align*}
(4) & \quad \text{a. I rode a } \textbf{unicorn} \text{ yesterday.} \\
& \quad \text{b. I rode a } \textbf{horse} \text{ yesterday.} \\
& \quad \text{English (Givón 1984: 438)}
\end{align*}

We see in (3) and (4) that even though the sentences have a different interpretation with respect to their truth value in the real world (the referents in (3a) and (4a) do not exist, as opposed to the ones in (3b) and (4b)), the English language does not seem to treat them differently. Both referents

\footnote{Gundel et al. (1993) call them giveness statuses.}


\footnote{In this study, I use this notion of discourse universe. Other related notions and concepts are “common ground”, used in Krifka & Musan (e.g. 2012b) 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.}
in (3) are expressed by the definite article, while both nouns in (4) occur with the indefinite article. As Givón (1984: 438) 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)

3.1.2 The link between linguistic expressions and referent types

Discourse referents are thus abstract elements of the discourse universe which in turn is part of the intersection of the speaker’s and the hearer’s mental spaces, but they nevertheless need to be linked to linguistic expressions from the utterances that represent the discourse. Since articles are linguistic expressions that are used to encode certain referent types, the nature of this link is important. This section sketches a way to define the link between referents and referring linguistic expressions. Since it would surpass the purposes of the present study, the goal of this section is not to offer a cognitively adequate model. It is intended as an explicit proposal of how referential functions, i.e. referent types, can be linked to linguistic expressions. “Referring expression” can be understood as any type of linguistic unit that can make a referring expression (e.g. phrases, article-noun combinations, larger event-encoding clauses, etc.).

Referring expressions are defined as an object with certain properties that can be represented in a typed feature structure. Feature structures are used in different linguistic frameworks such as Head-driven Phrase Structure Grammar or Construction Grammar. Without directly imposing any theoretical assumptions, they have the advantage of allowing a formal and explicit notation of relations between abstract types, their features, and their values. Feature structures are recursive and the values of features can be types again, having their own feature structure. For the purposes of the present study, referring expressions can be defined as having the type shown in Figure 3.2 with referentiality (ref) as one of its features. At the same time, referentiality takes an object of the type referent. The latter contains information of the referential function of the referring expression.

---

6For an overview on feature structures, see (Müller 2010: chapter 5).
Figure 3.2: Structure of a referring expression

An object of type \textit{referring.expression} has the following features: \textit{PHON}, \textit{SEM}, \textit{SYN}, and \textit{REF}. The feature \textit{PHON} takes a phonological object (\textit{phon-object}) as its value and specifies the phonological form of the word. The feature \textit{SYN} represents syntactic properties, e.g. case, but also the presence of modifiers or articles in the noun phrase could be included here. The value of \textit{SEM} corresponds to an object of the type \textit{sem-object}, specifying the context-independent lexical information of the type \textit{referring.expression} and other semantic properties. In that sense, the value for \textit{LEX} corresponds to an ungrounded “concept” as defined in Langacker (2008: 33f); we can imagine it as its denotation or lexicon entry. The feature \textit{REF} takes an object of type \textit{referent} as its value, which contains information on the referential functions of the referring expression. This is how the referents as elements of the discourse universe as defined in the previous section are built into the properties of a referring expression. Thus, the “link” between expressions and referents can be modelled as the \textit{REF} feature of an object of type \textit{referring.expression} which in turn takes an object of type \textit{referent} as its value.

A discourse referent of type \textit{referent} is characterized by the features \textit{REFERENT.TYPE} and \textit{CONCEPT}. The value of \textit{REFERENT.TYPE} defines the referent type as a \textit{ref.type-object}. I define different referent types in section 3.2; they are important since they serve as the basis for the distinction of different article types (cf. section 3.3). The other feature of the type \textit{referent} is \textit{CONCEPT}, taking an object of type \textit{concept-object} as its value.

An object of type \textit{concept-object} contains information about the context-independent, lexical meaning in its feature \textit{LEX}, which is copied from the \textit{sem-object}. The feature \textit{KIND} of the \textit{concept-object} refers to all referents that belong to the same kind as the referent of the expression in question. In other words, the value of \textit{KIND} can be defined as a set, namely the kind set \( \mathbb{K} \), which
contains all referents that belong to the concept of the referring expression. Figure 3.2 indicates this with the set notation as $\mathbb{K} := \{R^1, R^2, R^3, \ldots, R^n\}$; declaring $\mathbb{K}$ to be the set that contains the referents $R^1$ to $R^n$ of the concept belonging to the referring expression. This means that each referent, being of type referent, of any referring expression of type referring.expression contains the relation to all (potential) referents that share the same concept. While the implementation of the kind set $\mathbb{K}$ into the feature structure of a referring expression of the type referring.expression may not seem motivated at this point, I show in section 3.2 that the definition of the nonspecific and generic referent types relies on the concept of a kind set $\mathbb{K}$. Therefore, the latter is included into the properties of the referent, corresponding to an object of type referent in this feature structure representation. Nonspecific referents will be defined as any referent from the kind set $\mathbb{K}$ of its concept. Generic referents, on the other hand, will be defined as the kind set $\mathbb{K}$ itself and not as single referents thereof.

### 3.1.3 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, namely definite and specific referents.

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 $M_S$ and $M_H$. 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 $\mathbb{D}$.

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). Thus, a definition of identifiability based on (Sperber & Wilson 1986: 39) is presented below:

\[(5) \text{ Identifiability} \]

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 section 3.1.1, we can define identifiable referents and mutually identifiable referents as follows:

(6) **Identifiable referent**
A referent is identifiable by the speaker if the referent is an element of her mental space $M_S$, and the referent is identifiable by the hearer if the referent is an element of her mental space $M_H$.

(7) **Mutually identifiable referent**
A referent is mutually identifiable by the speaker and the hearer if the referent is an element of both of the speaker’s mental $M_S$ and the hearer’s mental space $M_H$.

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 2012a, 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 labelled “accommodation” (Karttunen 1974, Krifka & Musan 2012a, Lewis 1979, Stalnaker 1974). It corresponds to a repair strategy to rescue the interpretation of an utterance. Hawkins (1991: 413) illustrates hearer accommodation using the following example:

(8) **Context**:
Suppose Mary and Peter are looking at a landscape where she has noticed a distant church.

M: I’ve been inside that church.

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.

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, identifiability cannot only be marked by a linguistic expression, but that it can also be established through the use of a certain linguistic expression.

Another property of the concept of identifiability is that it is context-sensitive, which means that depending on the context, the speaker’s and 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. It is in this vein that 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 (9), the referent $R_{\text{boy}}$ of the expression $a \, \text{boy}$ is linked to a visible object (‘a boy’) in the discourse situation.

\begin{example}
(9) Context: \textit{S and H are on a ship in the ocean, looking at the water, when S spots a boy in the water.}

S: Look! There is \textbf{a boy} in the water! (Pirates of the Caribbean 1, movie)
\end{example}

Nevertheless, as the expression referring to $R_{\text{boy}}$ includes the indefinite article $a$, the referent $R_{\text{boy}}$ is marked as nonidentifiable, 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 as concepts to explain definiteness. 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 \textit{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 has been proposed in Frege (1892) and Strawson (1950). They integrated uniqueness as a presupposition for the felicitous use of the definite description (as 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

\footnote{Note that I treat referents and objects in the real world as different elements on different levels. Thus, strictly speaking, a referent can never be perceived directly, it can only be linked to an object from the world that is visible (cf. section 3.2).}
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, so that there must be 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, the familiarity approach which goes back to Christophersen (1939). He points out that instead of uniqueness, rather familiarity (or non-ambiguity of) with 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 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 (10):

(10) 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 (10) are newly introduced to 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\(^8\) 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.

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 Christoffersen (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. (Christoffersen 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. Therefore, uniqueness in the sense of an unambiguous relation and identifiability are equivalent concepts which I use in section 3.2.1 to define definite referents.

Indefiniteness, even more than definiteness, has been discussed strictly in combination with a specific linguistic form, the indefinite article \(a\) in English in most cases. Heim (1991: 32), for instance, argued that the indefinite article, in addition to the existence presupposition which it

---

\(^8\)Roberts (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.
shares with the definite article, has a non-uniqueness condition. She bases that on examples such as (11), for which she argues that the indefinite article is not felicitous because it is known that there is only be single referent.

(11)  a. I interviewed the (a) father of the victim.
 b. The (a) weight of our tent is under 4 lbs. (Heim 1991: 32)

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 (11) above. Note that the examples in (11) can also be accounted for by a more elaborate distinction of semantic noun types 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, accounting for the incompatibility with the indefinite article. While Löbner can account for the infelicitous use of the indefinite articles in examples like (11) making use of uniqueness being inherently built into the semantic properties of the referents, other scholars have argued against uniqueness being relevant for the use of the indefinite article.

Hawkins (1991: 420ff.) shows that the indefinite article can implicate uniqueness in some contexts, whereas in others, it cannot. Examples of these two types of contexts are shown in (12):

(12)  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: 417ff.)

In (12a), the referents of pizza and cake are marked as nonidentifiable by the indefinite article. The latter is required by the existential construction and the referents are understood as contextually unique. Example (12b), on the other hand, provides evidence for nonidentifiable 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 (13), 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).

(13)  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. More importantly, however, is that neither Heim’s and Hawkins’s approaches account for indefiniteness as a referential function. 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 should not be viewed as a primitive referential function but as the union of two referent types, namely specific and nonspecific referents.

### 3.2 Defining the referent types

This section defines the referent types that are necessary to distinguish between different article types. We can divide the referent types into the following four main types: definite, specific, nonspecific, and generic. As will be shown, the definite referent type is only an abstract referent type and needs to be divided further into subtypes of definite referents.

#### 3.2.1 Definite referents

As was already mentioned in the previous sections, definite referents are mutually identifiable by the speaker and the hearer i.e. they are elements of the intersection of the speaker’s mental space \( M_S \) and the hearer’s mental space \( M_H \). However, being elements in that set is not yet sufficient for a referent to be definite. Referents also need to be unambiguously identifiable in order to be definite, since it is their unambiguous identifiability that guarantees that they are the only salient referent of their kind.\(^9\) 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:

\begin{equation}
(14) \quad \text{Q-principle:}
\end{equation}

- **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

\[^9\]In the remainder of this study, I use “identifiable” and “identifiability” in the sense of mutual and unambiguous identifiability.
hand, can rely on the fact that when the speaker utters *That man is tall!*, then there is at least one and no more than one identifiable referent of the kind *man*.

In a way, the Q-principle is an informal and discourse-oriented version of the maximality condition\(^\text{10}\) in formal semantic approaches used for definite plurals (e.g. Schwarz 2012, Abbott 2010: 159-160). 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, plurals, instead of referring to a single contextually salient referent, refer to the maximality of contextually salient referents of a given kind. An utterance like *Those men are tall!* thus refers to all referents of kind ‘man’ that are contextually salient.

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

(15)  
a. [context: one bear]  
   The bear attacked.

b. [context: more than one bear]  
   The bears attacked.  

(Davis et al. 2014: 200)

(16)  
[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 (15) shows the successful application of definite singular and plural expressions. If there is only a single contextually salient referent as in (15a), 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 (15b), on the other hand, the use of the plural definite *the bears* to refer to more than one contextually salient referents of the kind ‘bear’ has to be interpreted as referring to all contextually salient referents of the kind ‘bear’. Example (16) 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’.

Uniqueness applied to singular referents is only a special case of maximality: it also refers to all referents of a given kind that are contextually salient, only that “all” means a single referent. Link (1998: 184) mentions 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

---

\(^{10}\) Another variant of this condition is the concept of “Inclusiveness”, proposed by Hawkins (1978: 157-167).
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.

Returning to the definition of a definite referent, we can formulate it using unambiguous identifiability being understood in terms of maximality:

(17)  **Definite referent** $R_{\text{def}}$
A definite referent is the maximal number of referents of a given kind that are mutually and unambiguously identifiable by the speaker and the hearer. A definite referent is therefore an element of the intersection of the speaker’s mental space $\mathbb{M}_S$ and the hearer’s mental space $\mathbb{M}_H$: $R_{\text{def}} \in (\mathbb{M}_S \cap \mathbb{M}_H)$

Given the definition of definite referents in (17), the next step is to make explicit how referents can become mutually and unambiguously identifiable by the speaker and the hearer. In the following sections, I introduce 7 relevant subtypes of definite referents that differ in how their mutual and unambiguous identifiability is achieved.

### 3.2.1.1 Spatial deictic referents

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”. An examples of a deictic referent is given in (18):

(18)  **Do you see the house** over there?

The presence of the object in the discourse situation makes the referent the 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:

(19)  **Deictic referent** $R_{\text{dei}}$
A referent is a deictic referent $R_{\text{dei}}$ if there is a function $f$ with its inverse function $f^{-1}$, and an object $O_{\text{dei}}$ in the discourse situation such that the following holds:

$$f(R_{\text{dei}}) = O_{\text{dei}} \land (f^{-1}(O_{\text{dei}}) = R_{\text{dei}}) \Rightarrow R_{\text{dei}} \in \mathbb{D} \Rightarrow R_{\text{dei}} \in (\mathbb{M}_S \cap \mathbb{M}_H)$$

The unambiguous link between the deictic referent $R_{\text{dei}}$ and the object $O_{\text{dei}}$ in the discourse situation is implemented by a function $f$ from the referent $R_{\text{dei}}$ to the object $O_{\text{dei}}$, and its inverse function, mapping the object $O_{\text{dei}}$ back to the referent $R_{\text{dei}}$. From that unique link between the referent $R_{\text{dei}}$ and the object in the discourse situation $O_{\text{dei}}$ follows that the referent $R_{\text{dei}}$ is an element of the discourse universe $\mathbb{D}$, which makes it an element of the intersection between the
speaker’s and the hearer’s mental spaces $M_S$ and $M_H$. Therefore, the deictic referent $R_{dei}$ is a definite referent.

### 3.2.1.2 Absolutely unique referents

There are two types of definite referents that are elements of the intersection of the speaker’s and the hearer’s mental spaces that are not identifiable based on anything related to the discourse situation itself. These are absolutely unique referents and recognitional referents. Absolutely unique referents are referents such as ‘sun’ or ‘moon’, which are 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:

(20) **The moon** is bright tonight.

Of course, this is not to say that we could not distinguish between different moons when talking about astronomy. 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 not compatible with nouns referring to these concepts; and hence, this referent type has to be distinguished in order to be treated separately. Absolutely unique referents entail that the kind set $K$ of an absolutely unique referent $R_{abs,u}$ only consists of a single referent. 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.

(21) **Absolutely unique referent** $R_{abs,u}$

A referent is an absolutely unique referent $R_{abs,u}$ if it is the only element in its kind set $K_{abs,u}$ such that the following holds:

$K_{abs,u} = \{R_{abs,u}\} \Rightarrow R_{abs,u} \in (M_S \cap M_H)$

---

11 A 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. While this classification of nouns accounts in an elegant way for certain effects in German and English with respect to the use of the definite article, the only relevant effects with respect to the use of articles in the languages in the world in general seems to be with two subtypes of what Löbner treats as individual nouns, namely absolutely unique nouns and proper nouns. Therefore, I do not distinguish between the other noun types here. As for the compatibility of definite articles with unique nouns and proper nouns, I show in sections 4.1.4 and 4.1.5 that the behaviour of definite articles in the world’s languages varies.
In this case, the absolutely unique referent $R_{abs,u}$ is unambiguously identifiable by both the speaker and the hearer independently from the discourse situation. Still, since both can identify the referent, it is part of both of their mental spaces $M_S$ and $M_H$ and therefore an element in the intersection of the two mental space sets, which makes it a definite referent. Note that this is why the discourse universe $D$ was defined as a subset of the intersection between the speaker’s and the hearer’s mental spaces in section 3.1.1: the speaker and the hearer can mutually and unambiguously identify referents outside of the discourse universe, as is the case for absolutely unique referents.

### 3.2.1.3 Recognitional referents

Recognitional referents represent another type of definite referents that are elements of the intersection between the speaker’s and the hearer’s mental spaces $M_S$ and $M_H$ without being an element of the discourse universe $D$. Recognitional referents are identifiable by both the speaker and the hearer because of previous shared experience or common knowledge.\(^{12}\) An example of a noun with a recognitional referent can be seen in (22). The referent *that dog* is identifiable and marked as such without having been mentioned previously in the discourse and without being linked to an object that would be physically present in the discourse situation.

(22) What happened to that dog (we used to have)?

The definition of recognitional referents is given below:

(23) **Recognitional referent $R_{recog}$**

A referent is a recognitional referent $R_{recog}$ if it is not the only element in its kind set $K_{recog}$ such that the following holds:

$$ (R_{recog} \in (M_S \cap M_H)) \land (R_{recog} \notin D) $$

Firstly, this definition requires the kind set $K_{recog}$ of a recognitional referent $R_{recog}$ to contain more than one referent so that we do not deal with an absolutely unique referent. Secondly, the definition requires the recognitional referent to be mutually and unambiguously identifiable by both the speaker and the hearer without being part of the discourse universe $D$. This ensures that we do not deal with a deictic or an anaphoric referent.

\(^{12}\)Recognitional referents were distinguished as a separate important subtype of definite referents most prominently in Himmelmann (1997: 61-82), who discusses them in German using the label “anamnestisch”.

59
3.2.1.4 Anaphoric referents

A very important and widely discussed subtype of definite referents 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 use. Another approach to definiteness and identifiability in which anaphoric uses play a major role is found in Ariel (1988, 1990, 2001). She examined the correlation between the linguistic expression of the discourse referent (NP, pronoun, demonstrative, etc.) and the distance between antecedent and 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. 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 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:

\[ R_{ana} = \text{Anaphoric referent} \]

A referent is an anaphoric referent $R_{ana}$ if there is a referent $R_{ante}$ such that the following holds:

\[ (R_{ana} = R_{ante}) \land (R_{ante} \in D) \Rightarrow R_{ana} \in D \]

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 $D$ 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.

3.2.1.5 Bridging referents

Another type of referents similar to anaphoric ones are bridging referents. “Bridging”\footnote{The 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), “inferrables” in Prince (1981), “accommodation” in Heim (1988), and as “indirect anaphora” in Erkül & Gundel (1987).} occurs when a nominal expression introduces a new referent $R_{brid}$, which is nevertheless licensed as def-
inite by a special relation to a previously mentioned referent \( R_{ante} \). It is not a proper anaphoric relation because the referents are not identical, but they are nevertheless linked in an unambiguous way. As Schwarz (2013: 536) notes:

One could consider them, for example, to be a special case of the anaphoric use, except that the antecedent is not the referent of the definite itself, but stands in some salient relationship to it. (Schwarz 2013: 536)

Hawkins makes the nature of this link more explicit:

[…]\ the mention of one NP, e.g. a wedding, can conjure up a whole set of associations for the hearer which permit the bride, the bridesmaids, etc. I shall refer to the first NP as ‘the trigger’, since it triggers off the associations, and to first-mention definite descriptions which are dependent on this trigger as ‘the associates’. (Hawkins 1978: 123)

The examples in (25) show typical bridging contexts. In (25a), the first part of the utterance refers to a book, establishing the referent of \( book (R_{ante}) \) as a part of the speaker’s and the hearer’s mental spaces \( M_S \) and \( M_H \). In the following part of the utterance, the referent \( R_{brid} \) of the word \( author \) is licensed as definite, because it has an unambiguous link to the already established referent of \( book (R_{ante}) \), which triggers the use of the definite article.


\[25a\]

b. John was driving down the street. \textbf{The} steering wheel was cold.  (Schwarz 2009: 6)

This unambiguous link between the two referents, which is not identity, usually involves a part–whole relation, product–producer relation, etc. Frame semantics (cf. Fillmore 1975, 1976) is inherently suitable 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)

We can apply this to examples like (25a) as follows: when a \( book \) is uttered, the frame of its referent \( R_{ante} \) evokes the concept of ‘author’ based on our world knowledge that books are written by people and these people are authors. By default, lacking more specific context, we can assume that each book has a single author. Therefore, the semantic frame of the concept ‘book’ includes one position for the concept ‘author’. This unambiguously links the referent \( R_{brid} \) of \textit{the author} to the referent \( R_{ante} \) of \textit{a book} and requires the noun \textit{book} to occur with a definite article, marking its
referent as definite, even though it is mentioned for the first time in the discourse. Figure 3.3 is an illustration of this unambiguous link between the two referents $R_{ante}$ of book and $R_{brid}$ of author enabled by the frame of the concept ‘book’ with one ‘author’ role. Imagining the kind set $\mathbb{K}$ of $R_{ante}$ that contains all potential referents of the concept ‘book’, each referent of the kind ‘book’ evokes a different referent of the kind ‘author’:

![Diagram](image)

Figure 3.3: Link between the referents of the kinds ‘book’ and ‘author’.

We can thus define a bridging referent as follows:

(26) **Bridging referent $R_{brid}$**

A referent is a bridging referent $R_{brid}$ if there is a function $f$ with its inverse function $f^{-1}$ and a referent $R_{ante}$ such that the following holds:

$$(f(R_{ante}) = R_{brid}) \land (f^{-1}(R_{brid}) = R_{ante}) \land (R_{ante} \in \mathbb{D}) \Rightarrow R_{brid} \in \mathbb{D}$$

This means that bridging referents $R_{brid}$ are unambiguously linked to a referent $R_{ante}$ that was uttered in the discourse situation at an earlier point in time and is thus an element in the discourse universe $\mathbb{D}$. The link between the preceding referent $R_{ante}$ and the bridging referent $R_{brid}$, however, is more complex than the identity function between anaphoric referents $R_{ana}$ and their preceding referents $R_{ante}$. For bridging referents $R_{brid}$, the link to the preceding referent $R_{ante}$ consists of a function $f$ from $R_{ante}$ to $R_{brid}$ and its inverse function $f^{-1}$ that returns $R_{ante}$ for $R_{brid}$. This means that the two referents do not share their identities, but there is some sort of unambiguous link from either of the two to the other one. The nature of the link itself is not important here and therefore simply expressed by an anonymous function $f$ with its inverse function $f^{-1}$; the important point is that these two functions make the link between a preceding referent $R_{ante}$ and the bridging referent $R_{brid}$ unambiguous, which in turn allows both the speaker and the hearer to unambiguously identify the bridging referent as well. Coming back to example (25a), one could ask why there is an function $f^{-1}$ that takes the referent $R_{author}$ of the author as an argument and returns the referent $R_{book}$ of a book, and that can be identified by the speaker and the hearer: based on our world knowledge, it is not necessarily the case that the author only wrote a single book. This is also hinted at in Figure 3.3, which only shows arrows from the different referents of type ‘book’ to the referents of type ‘author’, but not in the other direction. However,
since the immediately preceding discourse context features the utterance *a book*, there is only a single referent of the kind ‘book’ that is contextually salient. Therefore, we can assume that the link between the preceding referent $R_{book}$ and the bridging referent $R_{author}$ indeed consists of a function $f$ mapping the $R_{book}$ to $R_{author}$, as well as its inverse function $f^{-1}$, mapping the referent $R_{author}$ back to the same referent $R_{book}$ of kind ‘book’.

### 3.2.1.6 Contextually unique referents

Another type of definite referents is a contextually unique referent, which is very similar to bridging referents. What I call contextually unique referents in opposition to absolutely unique referents is what Hawkins (1978: 115-130) discusses as “larger situation use”. Contextually unique referents are mutually and unambiguously identifiable because they have an unambiguous link to the discourse situation that the speaker and the hearer are situated in. This link may be very narrow and concern objects or concepts directly related to the discourse situation, and it can involve referents that are linked to either the speaker or the hearer directly. Or, the link may be to a larger context, e.g. the city, or the country the discourse situation is placed in. Two examples of contextually unique referents are given in (27) and (28) below:

(27)  *S and H are in the small village H is from.*

    S:  Where is the church in this village?

(28)  *H is showing S her new flat.*

    S:  Where is the kitchen?

In example (27), the referent $R_{cont,u}$ of *church* is expressed by a noun with the definite article. The latter is used to indicate that the referent is contextually unique. This means it is the only salient referent of the kind $K$ linked to the concept of ‘church’ within the discourse situation. Because it is contained in the discourse situation, we can say that the discourse universe $\mathcal{D}$ also contains a referent of the concept ‘church’. Hence, contextually unique referents $R_{cont,u}$ are identifiable due to their unambiguous link to another referent $R_{ds}$ (ds for discourse situation) that is mutually and unambiguously identifiable based on the context of the discourse situation. Figure 3.4 sketches this link between $R_{ds}$ of the noun *village* and $R_{cont,u}$ of *church* from (27):
The relation between the referents in Figure 3.4 and Figure 3.3 are not identical by chance; contextually unique referents $R_{cont,u}$ are very similar to bridging referent $R_{brid}$ in that the referent is identifiable based on an unambiguously identifiable link to another referent that is mutually and unambiguously identifiable based on the discourse situation, only that the preceding referent is not necessarily mentioned in the discourse; it can be implicit. Thus, contextually unique referents can be defined as follows:

\[ (f(R_{ds}) = R_{cont,u}) \land (f^{-1}(R_{cont,u}) = R_{ds}) \land (R_{ds} \in \mathbb{D}) \Rightarrow R_{cont,u} \in \mathbb{D} \]

As a bridging referent, a contextually unique referent $R_{cont,u}$ is mutually and unambiguously identifiable because of its unambiguous link to another referent $R_{ds}$ which is an element of the discourse universe $\mathbb{D}$, and which is thus part of both the speaker’s and the hearer’s mental spaces. Due to the unambiguous link between these two referents, the contextually unique referent $R_{cont,u}$ is also an element of the discourse universe $\mathbb{D}$, and hence mutually and unambiguously identifiable by the speaker and the hearer as well.

The difference between bridging referents $R_{brid}$ and contextually unique referents $R_{cont,u}$ only lies in the properties that make the referents $R_{ante}$ and $R_{ds}$ they are linked to identifiable. As was shown for bridging referents, the linked referent is a preceding referent $R_{ante}$, uttered at an earlier point in time in the discourse situation. Contextually unique referents, on the other hand, are unambiguously linked to an identifiable but implicit referent of the discourse situation $R_{ds}$.

### 3.2.1.7 Establishing referents

What the preceding types of definite referents had in common is that their definite status was licensed by an unambiguous link to a referent that was already identifiable by both the speaker and the hearer. Therefore, we could say that the referent type licensed the use of the definite article.
Establishing referents are different. In contrast to previously discussed types of definite referents, the identifiability of this type is established. Establishing referents correspond to cases in which a referent is marked as definite, although it is not unambiguously identifiable by the hearer yet. Epstein (2002) argues from a cognitive perspective 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 speaker as well or will be shortly after the time of utterance. Two examples of establishing referents are shown in (30) and (31):

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

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

The referent \(R_{est}\) of rumour in (30) is not yet identifiable by the hearer at the point in time when the rumour is uttered, but it will be with the utterance of the complement clause that provides sufficient information on the referent \(R_{est}\). The referent is marked as definite so that the hearer can expect and establish identifiability. Establishing referents can be defined as follows:

\[
(32) \text{Establishing referent } R_{est} \quad A \text{ referent is an establishing referent } R_{est} \text{ at the point in time } t^i \text{ of the discourse situation, if } R_{est} \text{ is an element of the speaker’s mental space } \mathbb{M}_S \text{ in } t^i \text{ and if } R_{est} \text{ is an element of the hearer’s mental space } \mathbb{M}_{H} \text{ at a later point in time } t^{i+1} \text{ of the discourse situation such that the following holds:} \\
(R_{est}^i \in \mathbb{M}_S) \land (R_{est}^{i+1} = R_{est}^{i+1}) \land (R_{est}^{i+1} \in \mathbb{D})
\]

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 definite in order to signal to the hearer that she should treat the referent as an element of the discourse universe \(\mathbb{D}\) when it is uttered at \(t^i\), because it

---

14 This term is taken from Hawkins (1978).
will be identifiable and thus an element of the discourse universe \( D \) in the immediately following discourse segment at \( t^{i+1} \).

### 3.2.2 Specific referents

I define specific referents as single and particular referents from their kind set. In contrast to definite referents, at least the hearer is not able to identify the referent, and only knows that the linguistic expression is linked to a particular referent of a given kind.

This is similar but not identical to what is often understood as specificity, namely that the speaker is able to identify the referent but that the hearer is not.\(^{15}\) In addition, the concept of specificity is used in contrast with nonspecificity, defined as a discourse referent that neither the speaker nor the hearer can identify and that does not have any existence requirement (e.g. Baker (1973), Fillmore (1967), Karttunen (1976), Hawkins 1978: 203f, Anderson (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>
<thead>
<tr>
<th>Table 3.1: Traditional view of specificity and nonspecificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>identifiable by</td>
</tr>
<tr>
<td>speaker specific</td>
</tr>
<tr>
<td>hearer specific</td>
</tr>
</tbody>
</table>

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:

(33) Sarah wanted to talk to a colleague of mine, but I forgot which one.

(based on Ionin (2006: 182))\(^{16}\)

In (33), 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 that there is a particular referent that a colleague

\(^{15}\)As von Heusinger (2002: 247) puts it: “[...] I discuss the widely used description of specific NPs as a subclass of indefinite NPs as ‘known/identifiable to the speaker’ and show that this description is inadequate.”

\(^{16}\)The original example in Ionin (2006: 182) is: Sarah wants to talk to a colleague of mine, but I don’t know which one.
of mine is linked to, even though neither of them can identify the referent. Therefore, a specific referent $R_{spec}$ is a single referent of the kind set $\mathbb{K}$. While neither the speaker nor the hearer can identify the referent, they can assume that Sarah can identify the referent and that it could not be any referent of the kind ‘colleague of mine’. Figure 3.5 sketches this scenario:

![Diagram](image)

Figure 3.5: Anchoring of a specific referent

Figure 3.5 presents the referents of the kind set $\mathbb{K}$ of a colleague of mine on the left side. Both the speaker and the hearer know that one particular referent of these is unambiguously linked to the referent of Sarah, because Sarah can identify the referent of a colleague of mine. The speaker and the hearer, however, cannot unambiguously identify the latter referent directly, which is represented by the dotted lines from between the identifiable referent of the kind ‘Sarah’ and the ones of the kind ‘colleague of mine’. We can hence define a specific referent $R_{spec}$ as follows:

(34) **Specific referent $R_{spec}$**

A referent is a specific referent $R_{spec}$ if there is a function $f$ and a referent $R_{ds}$ such that the following holds:

$$f(R_{spec}) = R_{ds} \land (R_{ds} \in \mathbb{D}).$$

The link between the specific referent and the other identifiable referent in the discourse situation is again implemented as a function $f$ from the specific referent $R_{spec}$ to the identifiable referent $R_{ds}$ (the referent of Sarah in (33)) which is an element of the discourse universe $\mathbb{D}$. In contrast to bridging or contextually unique referents, this link is not unambiguous; for this reason, there is no identifiable inverse function $f^{-1}$, that would map the identifiable referent $R_{ds}$ back to the specific referent $R_{spec}$. In other words, in (33) both the speaker and the hearer know that the referent $R_{colleague}$ of a colleague of mine is unambiguously linked to the mutually and unambiguously identifiable referent $R_{sarah}$ of Sarah. Hence the function taking $R_{colleague}$ as an argument and mapping it to $R_{sarah}$ makes $R_{colleague}$ a particular one of its kind set $\mathbb{K}_{colleague}$. That neither the speaker can identify which referent $R_{colleague}$ of type ‘colleague of mine’ is linked to the expression colleague of mine is represented by the lack of an identifiable inverse function $f^{-1}$ that would take the referent $R_{colleague}$ as an input and map it to $R_{sarah}$. 

67
3.2.3 Nonspecific referents

While nonspecific referents are single referents from their kind set as well, they fundamentally differ from definite and specific referents in that nonspecific referents do not correspond to any particular referent from the kind set. An example of a nonspecific referent is given in (35):

(35) Lea wants to buy a bike, but she doesn’t know if she will find one.

The existence of a referent is often taken as the basis for the definition of nonspecific referents. Thus, the latter are usually assumed to lack an existence presupposition. 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, a “placeholder referent”\(^{17}\) from its kind set \( \mathcal{K} \), but no particular referent is linked to the expression used. With regard to the referent of a bike in (35), this scenario can be represented as follows:

![Diagram of nonspecific referent with its concept](image)

Figure 3.6: Nonspecific referent with its concept

Figure 3.6 shows the concept ‘bike’ which is linked to all referents that belong to this concept (the kind set). Since the expression a bike in (35) has to be linked to one of these referents, but since there is no information about the identity of the referent from the kind set that is linked, the referring expression can be linked to any of them, and it is either not important or not known which one it is. A nonspecific referent can thus be defined as a different type of referent, \( R_{nspec} \) which corresponds to a placeholder for any referent (\( R^1 \) to \( R^n \)) in its kind set \( \mathcal{K}_{nspec} \).

(36) **Nonspecific referent** \( R_{nspec} \)

A referent is a nonspecific referent \( R_{nspec} \) if there is a referent \( R_{nspec} \) with its kind set \( \mathcal{K}_{nspec} \) such that the following holds: \( R_{nspec} \in \mathcal{K}_{nspec} \)

\(^{17}\)Similarly 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 the kind set \( \mathcal{K} \).
3.2.4 Generic referents

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). In this study, I only treat the reference to kinds, as in the sentences in (37) below:

(37)  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.

I will take these expressions as referring expressions that are linked to a referent following Givón (1984: 440):

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)

Generic referents are of a different type than definite, specific, and nonspecific referents. They do not correspond to a single element of the kind set, but to all referents in this set. Thus, generic referents $R_{gen}$ can be treated to correspond to the kind set itself:\footnote{This is very similar to the description of generic referents in Chierchia (1998: 351): “So a kind can be manufactured out of a property by taking the largest member of its extension (at any given world).”}

(38) Generic referent $R_{gen}$

A referent is a generic referent $R_{gen}$ if there is a referent $R_{gen}$ and its kind set such that the following holds:

$R_{gen} = \mathbb{K}_{gen}$

3.3 Defining the article types

This section defines the 8 different article types that can be found in the world’s languages. Examples with a detailed discussion are provided in Chapters 4 to 6 for each article type. The 8 article types do not cover all of the four major referent types defined in the previous section. While we find different types of articles in the world’s languages that encode definite, specific, and nonspecific referents, articles whose main function it is to encode generic referents are not attested. This is why generic referents are not mentioned in the remainder of this section (for a discussion of generic referents and their coding by articles, see section 8.3.3.1).
3.3.1 Articles from the definite domain (definite, anaphoric, recognitional)

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 contexts as well. However, other uses of definite articles are not relevant for their definition, e.g., the occurrence with generic referents. Nor are definite articles required to encode all subtypes of definite referents. A definite article must be able to encode anaphoric (R_{ana}), bridging (R_{brid}), situational unique (R_{cont,u}), and establishing (R_{est}) referents. Deictic referents R_{dei} can be marked by the definite article in many but not all languages. We find definite articles with another diachronic source which are systematically used in the relevant definite contexts but do not occur in spatial deictic contexts due to their source element. A similar restriction concerns absolutely unique referents R_{abs,u} (cf. section 4.1.4). Definite articles in the languages of the world also differ with regard to their compatibility with this referent type. Since these articles are otherwise systematically used to encode different types of definite referents, I regard them as definite articles. Since the ability of definite articles to encode recognitional referents R_{recog} is not systematically documented in grammars, I do not include the use of a marker with recognitional referents as a criterion for definite articles for practical reasons. The same holds for establishing referents R_{est}. Following Himmelmann (1997: 93-101), definite articles are expected to encode both recognitional and establishing referents. He argues that both referent type play a central role in the development of definite articles. However, since this use is not well-documented for many definite articles in the sample, I do not include the coding of establishing referents as a necessary criterion either. This leaves us with the following definition of definite articles:

\[ (39) \quad \text{Definite article (art:def)} \]

A definite article is an article that systematically occurs with anaphoric and contextually unique referents. It may also occur with other types of definite referents. It does not occur with specific and nonspecific referents.

Anaphoric articles represent another type of articles from the definite domain. They are defined as follows:

\[ (40) \quad \text{Anaphoric article (art:ana)} \]

An anaphoric article is an article that systematically occurs with anaphoric referents. It does not occur with spatial deictic and contextually unique referents from the definite domain. It does not occur with specific and nonspecific referents either.

I treat such adnominal anaphoric markers as articles and not as demonstratives here (cf. section 7.1). This is because, in contrast to demonstratives, anaphoric articles have a less basic referential
function and are not universally found in the world’s languages (most languages use the demonstrative to encode anaphoric referents or the definite article, if available). Markers that encode both deictic and anaphoric referents are not treated as articles here but as demonstratives: the extension of deictic demonstratives to the marking of anaphoric referents is very common (if not universal) across the world’s languages.

The last article type from the definite domain is the recognitional article:

(41)  **Recognitional article** (\texttt{art:recog})

A recognitional article is an article that systematically occurs with recognitional referents. It does not occur with other types of referents.

A recognitional article is a marker whose only function it is to mark referents that are identifiable based on previous shared experience or knowledge. As is shown in section 4.3, we find such markers in the languages of the world. Similarly to anaphoric articles, I treat recognitional markers as articles, since they are referential markers which, unlike demonstratives, are not universal or crosslinguistically common (most languages use spatial demonstratives to mark recognitional referents).

### 3.3.2 Articles of the indefinite domain (exclusive-specific, nonspecific, indefinite)

In spite of the many treatments of indefinite articles over the course of the past decades, there is not much agreement on how they should be viewed, defined, or located in grammar. The problems associated with the definition, categorial status, and/or functions of indefinite articles are illustrated, for instance, by the way they are discussed in linguistic dictionaries. Rather than being of any real help, such “definitions” may be more of a burden to someone consulting these dictionaries. \cite{Heine:1997:67}

In the indefinite domain, we find three types of articles: exclusive-specific, nonspecific, and indefinite articles. Exclusive-specific articles can be defined as follows:

(42)  **Exclusive-specific article** (\texttt{art:exspec})

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

The nonspecific article can be viewed as the counterpart of the exclusive-specific article within the indefinite domain:
A nonspecific article is an article that systematically occurs with nonspecific referents. It does neither occur with definite referents nor with specific referents.

The third and probably most prominent type of articles from the indefinite domain is the indefinite article.

An indefinite article is an article that systematically occurs with specific and nonspecific referents. It does not occur with definite referents.

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 that I use to indicate that the article encodes a combination of referent types.

### 3.3.3 Domain-crossing articles (inclusive-specific, referential)

There are two more article types in the world’s languages that do not belong to either the definite or the indefinite domain; they occur with referents from both domains. These articles are inclusive-specific and referential articles.

Inclusive-specific articles can be defined as follows:

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

Inclusive-specific articles are thus the counterpart of exclusive-specific articles that are additionally used to mark definite referents. Again, expressing the different referent types should be understood as coexpression.

Referential articles are the last type of articles and they are the one with the largest functional domain:

A referential article is an article that systematically occurs with anaphoric, contextually unique, specific, and nonspecific articles. It may also occur with other types of definite referents.
In contrast to all other article types, referential articles do not mark a the referential function of the noun that they occur with; rather, they indicate that the noun is to be interpreted as a referential expression.

### 3.4 The relation between article types and referent types

This section addresses the question of the relation between referent types and article types in more detail, comparing the referent types proposed in section 3.2 with other previous hierarchies and scales concerning referentiality. Figure 9.1 gives an overview of the four major referent types to which the different article types can be mapped. The article types are: definite (def), recognitional (recog), anaphoric (ana), inclusive-specific (inspec), exclusive-specific (exspec), nonspecific (nspec), indefinite (indef), and referential (ref).

Figure 3.7: Overview of referent types and article types

Figure 9.1 sums up the relation between the article types and the major referent types of definite, specific, and nonspecific referents. In red, we see the article types from the definite domain, in blue those of the indefinite domain, and the articles in green are the ones that cross-cut this division into a definite and an indefinite domain.

While the linear order of the referent types as presented in Figure 9.1 seems to be chosen according to the functional domain of articles, they are in fact motivated by the properties of the referent types themselves as defined in section 3.2. All types of definite referents are unambiguously identifiable by both the speaker and the hearer. While the conditions that make the definite referents identifiable differ in their details, the unambiguous link was always modelled as a function $f$ mapping the referent to another referent or object that the speaker and the hearer can identify and its inverse function $f^{-1}$ mapping the latter back to the definite referent in question. Specific referents, on the other hand, were defined as referents that are linked by a function $f$ to
Crucially, there is no identifiable inverse function $f^{-1}$ that would take the identifiable referent as an argument and map it back to the specific referent. This is why specific referents are linked and particular, but they are not unambiguously identifiable. Moreover, this definition makes specific referents inherently a supertype of definite referents, or definite referent a special subtype of specific referents. Nonspecific referents, on the other hand, were defined as single elements of their kind sets, but without further information of their identity. Again, the definition of nonspecific referents only as elements in the kind set makes them a supertype of specific and definite referents. Only generic referents are not a supertype of all other referents, since they are not defined as single elements of the kind set, but as the kind set itself. Figure 3.8 shows this structure, which is entailed by the definitions of the four major referent types.

In a similar vein, we find hierarchies with respect to the structuring of referential functions in previous works. The three relevant hierarchies are the definiteness hierarchy in Croft (2003: 132) (based on the discussion in Comrie (1989: 132-35)), the givenness hierarchy in Gundel et al. (1993), and the reference hierarchy in Dryer (2014), going back to Givón 1978. What they have in common is that they structure and distinguish referential functions by mapping linguistic forms onto these functions. The three hierarchies are presented in (47) to (49).

(47) definite > specific > nonspecific  
     (definiteness hierarchy, Croft 2003)

(48) in focus > activated > familiar > uniquely identifiable > referential > type identifiable  
     (givenness hierarchy, Gundel et al. 1993)
The definiteness hierarchy shown in (47) is based on differential object marking in Spanish, Turkish, and Persian shown in Comrie (1989: 132-135). In Turkish and Persian, accusative marking is reserved to nouns with definite referents. Spanish marks patients in transitive contexts if they are animate referents and definite or specific; nonspecific referents are not marked. The hierarchy (based on an entirely different coding phenomenon) essentially shows a similar ranking of referential functions as the hierarchy in Figure 3.8, although it does not clearly distinguish between nonspecific and generic referents. The latter are characterized as “those identifiable only as a type, not as a specific instance or token” (Croft 2003: 132). Nevertheless, this hierarchy shows that also other phenomena in the grammar, namely differential object (argument) marking, reflect this alignment of definite, specific, and nonspecific functions.

The givenness hierarchy in Gundel et al. (1993: 275) structures referential functions in combination with discourse prominence. These functions are called “givenness statuses” and correspond to “six cognitive statuses relevant to the form of referring expressions in natural language discourse” (Gundel et al. 1993: 275). The statuses are not identical with the referent types as defined here, but some overlap: Gundel et al. (1993) distinguish between four types of definite referents (varying in discourse prominence), specific, and what they call generic referents, which also includes nonspecific referents. Gundel et al. (1993: 276) mention explicitly that the givenness hierarchy should not be understood as mutually exclusive but as an implicational hierarchy “such that each status entails (and is therefore included by) all lower statuses, but not vice versa”. The main problem of the givenness hierarchy is that it cannot separate the organization of referent types and the linguistic expressions; it is intended as a universal hierarchy of semantic functions whose definition is based on linguistic expressions in a single language, namely English. Also, the combination of referentiality and discourse prominence makes it very difficult to properly separate the statuses. Nevertheless, it is a hierarchy of referent types that cover the entire referential space.

The third relevant hierarchy is the reference hierarchy of Dryer (2014: 235) in (49). This hierarchy is very similar to the one proposed in the present study. “Pragmatically specific indefinites” largely correspond to establishing definites. This “hierarchy” 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 sin-
gle article can co-express different (also more than two) functions, but only as long as their are neighboring 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)

Thus, the motivation for this hierarchy is the organization of referential functions depending on the combinations of functions that are encoded by articles. Therefore, the hierarchy proposes more hierarchical distinctions in the definite domain, where I do not impose any hierarchical structure: Dryer (2014) distinguishes between higher anaphoric and lower nonanaphoric definite referents; predicting that there are articles that only code anaphoric definites, but also predicting articles that mark non-anaphoric referents. While anaphoric articles are widely attested, I show in section 8.3.3.2 that the existence of non-anaphoric articles in senso stricto in the languages of the world is rather questionable.

3.5 Summary

This chapter dealt with the definition of different article types based on their referential functions. In section 3.1, I introduced the relevant concepts related to referentiality, and defined four major referent types that correspond to referential functions in section 3.2: definite, specific, non-specific, and generic referents. Section 3.3 defined 8 article types based on the referent types that they express: definite, anaphoric, recognitional, exclusive-specific, nonspecific, indefinite, inclusive-specific, and referential articles. According to the referent types they express, we can distinguish three groups: articles in the definite domain, articles in the indefinite domain, and domain-crossing articles. Section 3.4 discussed the relation between the article types and referent types, proposing a hierarchy of referent types onto which articles can be mapped according to their functional domains. Table 3.2 on the next page summarizes the different referent types from the definite and indefinite domains, indicating the functional domains of each article type which corresponds to the expression of different combinations of article types.
Table 3.2: Summary of referent types and article types

<table>
<thead>
<tr>
<th>Referent Type</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&lt;sub&gt;dei&lt;/sub&gt;</td>
<td>Deictic referent identifiable based on its unambiguous link to an object in the discourse situation.</td>
<td>Do you see the house over there?</td>
</tr>
<tr>
<td>R&lt;sub&gt;recog&lt;/sub&gt;</td>
<td>Recognitional referent identifiable based on shared previous experience or (common) knowledge.</td>
<td>What happened to the dog (we used to have)?</td>
</tr>
<tr>
<td>R&lt;sub&gt;abs.u&lt;/sub&gt;</td>
<td>Absolutely unique referent identifiable because it is the only referent of its kind.</td>
<td>The Earth is round.</td>
</tr>
<tr>
<td>R&lt;sub&gt;ana&lt;/sub&gt;</td>
<td>Anaphoric referent identifiable based on shared identity with a previously mentioned referent.</td>
<td>A: Did you like the movie last night? B: The movie was great!</td>
</tr>
<tr>
<td>R&lt;sub&gt;brid&lt;/sub&gt;</td>
<td>Bridging referent identifiable based on its unambiguous link to another previously mentioned referent.</td>
<td>Eva bought a book. The author is French.</td>
</tr>
<tr>
<td>R&lt;sub&gt;cont.u&lt;/sub&gt;</td>
<td>Contextually unique referent identifiable because it is the only salient referent of its kind in the discourse situation.</td>
<td>What’s the best way to the center?</td>
</tr>
<tr>
<td>R&lt;sub&gt;est&lt;/sub&gt;</td>
<td>Establishing referent marked as identifiable and thus constructed as such.</td>
<td>Did you hear the news? They are going to close the museum.</td>
</tr>
<tr>
<td>R&lt;sub&gt;spec&lt;/sub&gt;</td>
<td>Specific referent is not identifiable but linked to a particular referent of its kind.</td>
<td>I met a strange new neighbour yesterday.</td>
</tr>
<tr>
<td>R&lt;sub&gt;n spec&lt;/sub&gt;</td>
<td>Nonspecific referent corresponds to a single, but no particular (any), referent of its kind.</td>
<td>Do you have a pen? Any pen will do.</td>
</tr>
<tr>
<td>R&lt;sub&gt;gen&lt;/sub&gt;</td>
<td>Generic referent corresponds to the set of all referents of its kind.</td>
<td>I like cats.</td>
</tr>
</tbody>
</table>
Chapter 4

Articles in the definite domain

This chapter gives an overview of definite (section 4.1), anaphoric (section 4.2), and recognitional articles (section 4.3) as defined in Chapter 3. 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 argue in section 4.1.3 that the coding of (spatial) deictic referents is not a necessary criterion for definite articles. Rather, the occurrence in deictic contexts of definite articles can be accounted by the fact that most definite articles originate from demonstratives that are used in such contexts. As I show for Kaqchikel, definite articles with sources other than 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 4.1.4) and with proper nouns (cf. section 4.1.5), showing that definite articles can but do not have to occur in such contexts, without any noticeable crosslinguistic trend.

4.1 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 and establishing referents as well. But as was mentioned in section 3.3.1, especially bridging and establishing contexts are not always well-documented, which results in the practically motivated decision to treat an article as definite on the basis of anaphoric and contextually unique referents only. This section first zooms in on the definite articles in Kaqchikel (Mayan, Guatemala) in section 4.1.1 and Mokpe (Bantu, Cameroon) in section 4.1.2. Definite articles with regard to the expression of deictic referents are discussed in section 4.1.1. The compatibility of definite articles with absolutely unique referents and proper nouns is then addressed in sections 4.1.4 and 4.1.5, respectively.
4.1.1 Kaqchikel

Kaqchikel (Mayan, Guatemala) has a marker *ri* that is usually treated as a definite article (R. M. Brown et al. 2006, Chonay Chonay 2006). In this section, I show that *ri* is systematically used to mark anaphoric, bridging, contextually unique, and establishing referents. Moreover, it does not occur with specific or nonspecific referents, for which Kaqchikel uses the indefinite article *jun*.

In (1), the article *ri* obligatorily marks an anaphoric referent in a short conversation: the expression *ixoq* ‘woman’, uttered by speaker B, refers back to the referent of *jun* *ixoq* uttered by speaker A.

(1) A: tak’ xinapon chuwe choch *+[jun *ixoq] ruk’in kayi aq’ala* [art.indef] when arrive.pst.1sg to.1sg house see.pst.1sg ART:INDEF woman with two child ech’okol chuhi qaq’
    seated close fire
    ‘When I came home, I saw a woman with two children sitting in my kitchen.’

B: awetaman ruwäch [*+(ri) *ixoq]? [art:def] know.pst.2sg eyes.3sg ART:DEF woman
    ‘Did you know the woman?’ Kaqchikel (prim. data)

Two similar anaphoric contexts in which the definite article occurs with the noun are illustrated in (2) and (3). In (2), B’s reply to A’s question makes use of the definite article *ri* with *wäy* ‘tortilla’ to refer back to the referent of the utterance made by A. Example (3) shows that the definite article is also obligatorily used with the abstract noun *ajob’ab’al* ‘love’ to mark anaphoric reference.

(2) A: ninwajo ntä wäy want.1sg eat tortilla
    ‘I would like to eat tortilla.’

B: [*+(ri) wäy] k’o chuchi qaq’ ART:DEF tortilla exist close fire
    ‘The tortilla is in the kitchen.’ Kaqchikel (prim. data)

(3) *ri* amistad xok ajab’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 (prim. data)

Another relevant and similar function of definite articles is the coding of bridging referents. Similarly to anaphoric referents, bridging referents are definite because they involve an unambiguous link between a preceding referent and the bridging referent. Examples (4), (5), and (6) illustrate the use of the definite article in Kaqchikel with bridging referents. In (4), we find 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
link a single referent of the concept ‘roof’. In (5) and (6), the reference to preceding k’ulbik ‘wedding’ and ch’ich’ ‘car/taxi’ enables the definite marking of xten nik’ule ‘bride’ and chofer ‘driver’ accordingly: the concepts of the preceding referents open a semantic frame with one position for the bridging referent, linking the latter to the preceding referent and making it definite.1 In all the three bridging examples shown below, the use of the definite article is obligatory.

(4) xk’atz’enkän jun jay tziaqchik chila pa tinamet. [*(ri) ruwi] jmul see.PAST.IPL 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 (prim. data)

(5) 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 (prim. data)

(6) 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 (prim. 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 it is linked to another identifiable referent from the discourse situation. Example (7) allows for the unambiguous identification of the referent of k’amor-b’äl ‘president’ because the concept of ‘president is unambiguously linked to the concept of ‘country’, which can be interpreted with the larger discourse situation being situated in Guatemala.2 Hence, the definite article is used.

(7) [*(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 (prim. data)

---

1 I use the notions “definite” and “identifiable” interchangeably.
2 Concepts such as ‘president’ correspond to functional nouns according to the classification of noun types in Löbner (1985). He notes: “For functional nouns, the relation that defines their reference is a function. Functions relate objects unambiguously (or one-to-one) to others” (Löbner 1985: 293). Thus, contextually unique referents as defined here and functional nouns are similar and compatible concepts, only that contextually unique referents are defined as referential functions independently from their linguistic expression, whereas functional nouns are defined as specific types of nouns in specific languages.
Another example of a contextually unique referent is presented in (8). 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, Patzun in Guatemala). Therefore, the referent of *aq’omanel-jay* ‘hospital’ is contextually unique and hence obligatorily marked by the definite article as well.

(8) ankuchi k’obi [*'(ri) aq’omanel-jay] chin re tinamet re?
   where EXIST ART:DEF cure-house in DEM town DEM
   ‘Where is the hospital in this town?’ Kaqchikel (prim. data)

Furthermore, definite articles are expected to mark establishing referents, i.e. mark a referent as identifiable so that the hearer can establish identifiability. An example of such context is a relative clause that follows the noun and expresses the referent which is established as definite. This is shown in examples (9) and (10). 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.

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

(10) [*'(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 (prim. data)

The previous examples thus showed that the definite article in Kaqchikel systematically occurs with anaphoric, bridging, contextually unique, and establishing referents. I will now show that *ri* is restricted to definite contexts and does not occur with nonidentifiable referents, which are marked by the indefinite article *jun* in Kaqchikel. Examples (11), (12), and (13) provide specific contexts and show that *ri* cannot be used with specific referents. Specific referents correspond to particular referents that are linked to the expression used, but they are not identifiable by all discourse participants. This is the case for the referents of *jay* ‘house’, *lugar* ‘place’, and *chkop* ‘animal’: the hearer cannot identify them, and they have to be marked by the indefinite article *jun*.

(11) xk’atz’enkăn [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 (prim. data)
Examples (14), (15), and (16) below show that the same holds for nonspecific referents. Nonspecific referents correspond to a single but not particular referent of its kind. In (14) for instance, any single referent of concept ‘pen’ can be referred to by the noun atzib’ab’al ‘pen’.

(14) 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 (prim. data)

(15) ninwajo [jun ch’ich’]
want.1sg ART:INDEF car
‘I want to have a (any) car.’ Kaqchikel (prim. data)

(16) nk’atzin chuwe [jun b’anoy-qutun]
need.1sg to.me ART:INDEF prepare-food
‘I need a (any) cook.’ Kaqchikel (prim. data)

Thus, we saw that ri in Kaqchikel is a definite article, occurring with different types of definite referents, and being absent in specific and nonspecific contexts.

4.1.2 Mokpe

Another example of a definite article can be found in Mokpe (Bantu, Cameroon). 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.

As is commonly found in Bantu, Mokpe has an elaborate gender system which is traditionally referred to as noun class system, since the values of gender and number of the noun occur as a fused marker on the noun, on other elements in the noun phrase and as agreement markers on the verb. Section 8.2.2 discusses the morphological properties of the definite article in Mokpe in more detail. Here, I give a brief overview of the exponents of the different classes and argue that they have to be exponents of a single definite article based on their distribution with nouns in definite contexts.

Table 4.1 shows the article exponents for each class. The class distinction is adapted from Atindogbe (2013); all examples and the analysis of the article system are based on my fieldwork.
on Mokpe. Atindogbe (2013: 11) mentions articles in Mokpe as follows: “the indefinite and definite articles are rendered by the low and the high tone respectively”. I argue that the definite article indeed surfaces as high tone on the noun. However, a noun that is not marked by this high tone does not correspond to a noun with an indefinite article. I treat it rather as the citation form which is used in all non-definite contexts. In addition, Atindogbe (2013: 11) does not mention the segmental marker as an article exponent. Either, it is simply left out in the description of Atindogbe (2013), or the variety described may differ from the Mokpe variety spoken by my informants.

Table 4.1: Mokpe article paradigm

<table>
<thead>
<tr>
<th>class</th>
<th>noun</th>
<th>article+noun singular</th>
<th>class</th>
<th>noun</th>
<th>article+noun plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mòtò</td>
<td>è mòtò</td>
<td>2</td>
<td>wàtò</td>
<td>wàtò</td>
</tr>
<tr>
<td></td>
<td>ɲmánà</td>
<td>è ɲmánà</td>
<td></td>
<td>wànà</td>
<td>wànà</td>
</tr>
<tr>
<td>3</td>
<td>móléli</td>
<td>móléli</td>
<td>4</td>
<td>méléli</td>
<td>méléli</td>
</tr>
<tr>
<td></td>
<td>mólèngò</td>
<td>mólèngò</td>
<td></td>
<td>mélèngò</td>
<td>mélèngò</td>
</tr>
<tr>
<td>5</td>
<td>likàla</td>
<td>likàla</td>
<td>6</td>
<td>màkàla</td>
<td>màkàla</td>
</tr>
<tr>
<td></td>
<td>likpà</td>
<td>likpà</td>
<td></td>
<td>màkpà</td>
<td>màkpà</td>
</tr>
<tr>
<td>7</td>
<td>ëtàngùlè</td>
<td>ëtàngùlè</td>
<td>8</td>
<td>ñétàngùlè</td>
<td>ñétàngùlè</td>
</tr>
<tr>
<td></td>
<td>ëlélé</td>
<td>ëlélé</td>
<td></td>
<td>ñélélè</td>
<td>ñélélè</td>
</tr>
<tr>
<td>9</td>
<td>mbèžà</td>
<td>è mbèžà</td>
<td>10</td>
<td>í mbèžà</td>
<td>í mbèžà</td>
</tr>
<tr>
<td></td>
<td>mbólì</td>
<td>è mbólì</td>
<td></td>
<td>í mbólì</td>
<td>í mbólì</td>
</tr>
<tr>
<td>14</td>
<td>wòzò</td>
<td>wòzò</td>
<td></td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>wǒngò</td>
<td>wǒngò</td>
<td></td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>19</td>
<td>inõní</td>
<td>inõní</td>
<td></td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

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. 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, 9, and 10, the article exponent is a segmental marker; è for class 1 and 9, and í for nouns of class 10. 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. This is a strong argument for the form-independence of articles. As we will see in this section, the tonal process has the same function as the segmental
marker è in that both markers occur in the same types of contexts with nouns. Therefore, I treat them as equal exponents of the definite article in Mokpe.

That the definite article is used in anaphoric contexts is shown in (17) and (18). In both examples, the referents of mólánà ‘woman’ and màlúwà ‘water’ are identifiable because they were previously mentioned in the discourse. The definite article is required in such contexts. In (17), we see the article exponent è with a noun of class 1. In contrast, the article surfaces as high tone with the noun màlúlwà ‘water’ which belongs to class 6.

(17) A: è bóná na dżêlì, n-êni mólánà à lîj-ô kîjêni nà bánà bá
ART:DEF time I came 1SG-see woman.CL1 CL1 sit-in kitchen with children.CL2 CL2 wáké
two
‘When I came (home), I saw a woman with two children sitting in the kitchen.’

B: ò mà wijá [è mólánà]?
you PST know ART:DEF woman.CL1
‘Did you know / recognize that woman?’

Mokpe (prim. data)

(18) A: nà bêja è pîzá màlúwá
I feel ART:DEF thirst water.CL6
‘I am thirsty.’

B: màlúwá má bél-ô kîjêni
water.CL6 ART:DEF CL6 is-in kitchen
‘(The) water is in the kitchen.’

A: dżä tè, nà zéni màlúwá/
come please I see.NEG water.CL6 ART:DEF water.CL6
‘Please come, I cannot find the water.’

Mokpe (prim. data)

The definite article also has to be used with nouns in bridging contexts in Mokpe. Examples (19) and (20) illustrate this use. In (19), the referent of mòtìlèlì ‘writer’ is identifiable because of its unambiguous link to the previously mentioned referent of kâtì ‘book’. Since the noun mòtìlèlì ‘writer’ belongs to class 1, the article surfaces as è.

(19) mòkákè àndí kâtì. [è mòtìlèlì] à bélì.ndí mòtà bëndì.
Mokake buy.3SG book ART:DEF writer.CL1 CL1 be person.AM French
‘Mokake bought a book; the author is French.’

Mokpe (prim. data)

Example (20) shows the Mokpe counterpart of example (4) from Kaqchikel with a part-whole relation between a house and its roof. Again, we see that the definite article is used together with the expression of likândò ‘roof’. Since this noun belongs to class 5, it requires the high tone exponent.
Example (21b) shows a third example for bridging contexts. By mentioning the referent of è mbówà jèní ‘his village’, the concept of ‘village’ makes it possible to unambiguously link the people living in that village, and the “typical” landscape surrounding a village (e.g. a river and mountains) to it. Therefore, the referents of wàtò ‘people’, mòlélí ‘food’, mòʒɔ̀ ‘river’, and βàkó ‘mountain’ are unambiguously identifiable and thus definite. We see that the referents need to be marked by the definite article. The nouns wàtò, mòlélí, and mòʒɔ̀ occur with a high tone on their first tone-bearing unit, while the noun βàkó takes the segmental article exponent è.

Similarly to bridging referents, a referent that is contextually unique requires the marking by a definite article. Example (41), the counterpart of (7) from Kaqchikel, illustrates the use of the definite article in such contexts. Uttered in Cameroon, the referent of ‘president’ is contextually unique in the country, which is why the referent of mòkànèlì ‘leader (president)’ is unambiguously identifiable and requires to be marked by the article. Moreover, the utterance in (41) contains the noun ndʒùmá ‘fight, strike’ which is equally marked as definite. Again, due to the location of the discourse situation being the anglophone part of Cameroon in 2017 (the city of Buea), both the speaker and the hearer know about the strikes against the repressions from the francophone government towards the anglophone population. Therefore, the referent of ndʒùmá ‘fight, strike’ is contextually unique and unambiguously identifiable by both the speaker and the hearer.

---

85
(22) [*[ê] mòkànlè] à ʒóβ-è ñé dʒóβ-è ndefµà
leader.cl1 NEG-have time for fight.cl9

literally: ‘The leader does not care about the strike.’

Mokpe (prim. data)

Examples (23) and (24) below show that the definite article in Mokpe has to be used with establishing referents as well. In those two examples, the referents marked as definite are made identifiable to the hearer only in the immediately following discourse segment.

(23) [*[ê] mò tô] Ngúɗú á mùŋmá wűnûwû à wëlè
man Ngudu 3sg meet.pst last.night 3sg call

Mokpe (prim. data)

(24) njëndì/ [njëndì] mó ñë kɔ̌wí ñàmà
news.cl3 ART:DEF animal catch

Mokpe (prim. data)

In order to qualify as a definite article, the article in Mokpe also needs to be absent in specific and nonspecific contexts. Examples (25) to (27) show specific referents which are expressed by the bare nouns ndáwò ‘house’, ámà ‘animal’, and lóbá ‘bundle/gift’ in (25), (26), and (27), respectively. That the definite article is absent with nonspecific referents as well is shown in (28) to (30).

(25) i mënè njëndì ndáwò en mbówà
see.pst.1pl small ugly house in village

Mokpe (prim. data)

(26) ñngg! ê njëndèli è kɔ̌wí njëmá
look.imp ART:DEF trap.cl9 catch animal
‘Look! The trap caught an animal.’

Mokpe (prim. data)

(27) nò óβélì lóbá
1sg.2sg have bundle
‘I have a gift for you.’

Mokpe (prim. data)

(28) ówélì èkí já li lë moβě mbówà
place of INF eat DEM:PROX village.q
‘Is there a place to eat in this village?’

Mokpe (prim. data)

(29) nà àŋrà mò tô wà mòkàw à wí lìzřŋò nòmà nà mà wíjè
I need person AM hunt 3sg know hunting like I pst know
‘I need a (any) hunter who can hunt like I used to do.’

Mokpe (prim. data)
Thus, we could establish the different markers in Mokpe as exponents of the definite article, used with nouns in the relevant definite but not indefinite contexts.

4.1.3 Definite articles with spatial deictic referents

Although there are definite articles that can encode spatial deictic referents, I argue in this section that marking deictic referents is not a necessary criterion for definite articles. The use of a definite article with deictic referents is probably often due to the development of definite articles from spatial demonstratives (or any marker that can be used with spatial deictic referents, for that matter). In such a scenario, the definite article simply retains the function of marking deictic referents as definite. This is a novel view insofar as deictic referents are often treated as the most typical definite referents, reflected by the fact that many expressions for definiteness develop from demonstratives (e.g. Lyons 1999: 322). In Kaqchikel, for instance, the definite article ri is not compatible with deictic referents while it occurs systematically with non-deictic definite referents as was shown in section 4.1.1. Therefore, I do not view the marking of deictic referents as a necessary criterion for definite articles. Instead, their compatibility with deictic referents may often simply be the result of their development from demonstratives that are also used to mark spatial deictic referents.

In Mokpe, the definite article is compatible with deictic referents. Examples (31) and (32) show its use in such contexts. Example (33) shows that the demonstrative marker and the definite article can be used interchangeably to refer to a deictic referent.

(31) Context: S points at the only table in the room.
S: [è tèwèlì] è gbèjáná wí wèwámbo
  ART:DEF table CL9 CL9 made of wood
  ‘The table is made of wood.’
  Mokpe (prim. data)

(32) òngɔ́! [è njùndèlì] è kòwí nàmà.
  look.imp ART:DEF trap CL9 CL9 catch animal
  ‘Look! The trap caught an animal.’
  Mokpe (prim. data)

(33) [ònò/ è ñmánà] á má kòkà liwòtèjá è βòndá nà mà ñmènè
  DEM:PROX ART:DEF child 3SG RES grown since ART:DEF time I PST see.o:3SG
  ‘This child has grown since the last time I saw her.’
  Mokpe (prim. data)

*I use “deictic” in the sense of “spatial deixis”.*
In Kaqchikel, on the other hand, the definite article \( ri \) does not occur in such deictic contexts. This is arguably due to its diachronic source, which is the anaphoric marker \( ri \ldots ri \). Thus, that the article is not compatible with deictic referents can be accounted for by the properties of the article’s source element which itself is not able to mark deictic referents. This is important because it makes the incompatibility of the definite article with deictic contexts only be a consequence of its source element rather than being an independent property of the definite article.

Examples (34a) and (34b) show two utterances with the referent \( ulew \) ‘ground’ marked by a demonstrative and the definite article, respectively. The referent of \( ulew \) in these sentences receives two different interpretations: with the demonstrative in (34a), it is necessarily interpreted as deictic, as is indicated by the translation. On the other hand, the noun \( ulew \) ‘ground’ with the definite article in (34b) only allows for a non-deictic referent; it cannot be accompanied by a pointing gesture and, without further context, corresponds to a generic statement.

\[
\begin{align*}
(34) & \quad a. \quad [la \ ulew] \ tz'il \\
& \quad \text{DEM ground dirty} \\
& \quad \text{‘The ground (here) is dirty.’} \\
& \quad b. \quad [ri \ ulew] \ tz'il \\
& \quad \text{ART.DEF ground dirty} \\
& \quad \text{‘The ground (in general) is dirty.’}
\end{align*}
\]

Kaqchikel (prim. data)

Examples (35) and (36) 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 (35a) and (36a) 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 (35b) and (36b), 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 (35b) and (36b) are only felicitous if the objects ‘sun’ and ‘moon’ are not directly visible neither to the speaker nor to the hearer.

\[
\begin{align*}
(35) & \quad a. \quad \text{man tatzu } [la \ q’ij] \\
& \quad \text{NEG look.IMP DEM sun} \\
& \quad \text{‘Don’t look at the sun.’ (the sun being directly visible, e.g. from outside of a house)} \\
& \quad b. \quad \text{man tatzu } [ri \ q’ij] \\
& \quad \text{NEG look.IMP ART:DEF sun} \\
& \quad \text{‘Don’t look at the sun.’ (the sun not being visible, e.g. from inside of a house)}
\end{align*}
\]

Kaqchikel (prim. data)
(36) a. \([\text{la } \text{ik’}] \text{ jani nqalaj chpam re aq’a re DEM moon very be.visible.3sg in DEM night DEM}\)

‘The moon is very bright tonight.’ (the moon not being visible, e.g. from inside of a house)

b. \([\text{ri } \text{ik’}] \text{ jani nqalaj chpam re aq’a re DEF.ART moon very be.visible.3sg in DEM night DEM}\)

‘The moon is very bright tonight.’ (the moon not being visible, e.g. from inside of a house)

As was already mentioned, it is plausible to assume that the definite article \(\text{ri}\) in Kaqchikel originates from an anaphoric marker \(\text{ri} \ldots \text{ri}\). The system of demonstratives including the anaphoric marker is given in Table 4.2 below (based on R. M. Brown et al. 2006: 26,149ff. and Chonay Chonay 2006: 103ff.):

<table>
<thead>
<tr>
<th>Demonstrative</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>proximal</td>
<td>re noun re</td>
</tr>
<tr>
<td>distal</td>
<td>la noun (la)</td>
</tr>
<tr>
<td>anaphoric</td>
<td>ri noun (ri)</td>
</tr>
</tbody>
</table>

For proximal \(\text{la} \ldots \text{la}\) and anaphoric \(\text{ri} \ldots \text{ri}\), the second postposed part is often omitted (the form \(\text{la}\) of the distal demonstrative used in (34a), (35a), and (36a) 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 \(\text{ri}\) corresponds to what I treat as the definite article \(\text{ri}\). Although \(\text{ri}\) was shown to occur with anaphoric referents in section 4.1.1, contexts that include anaphoric referents with the anteceding referent in the same sentence and uttered by the same speaker still require the full anaphoric form \(\text{ri} \ldots \text{ri}\) and do not allow for the short form \(\text{ri}\). Two examples to illustrate this are given in (37) and (38).

(37) a. \([\text{Lu } \text{xiroyoj jun b’eycik mambajota yich’o ruk’in [ri vinaq PROP Pedro call.pst.s:3sg.o:1sg one way want.neg.1sg talk with.3sg ANA person *(ri)] ANA}\)

‘Pedro called again. I don’t want to talk to that person.’ Kaqchikel (prim. data)
We can hence assume that the definite article *ri* originated from the anaphoric marker *ri …ri*, and did not have to undergo a functional extension from spatial deixis to anaphora. Instead, the emerging article started as an anaphoric marker, extending its function to the marking of establishing referents, contextually unique, and bridging referents. The fact that a definite article like *ri* in Kaqchikel does not extend its use to the marking of deictic definites can be accounted for by the presence of other demonstrative markers in the grammar. Demonstratives that cover the coding of deictic referents may result in the lack of functional pressure on the definite article to be used with deictic referents.

This example from Kaqchikel showed two important things. Firstly, definite articles do not necessarily encode deictic referents, even though many definite articles from the world’s languages may do so. Therefore, the coding of deictic referents cannot be taken as a criterion for definite articles. Secondly, the fact that many definite articles encode deictic definites rather follows from their origin as demonstrative (or similar) markers. If the source element of the definite article is not compatible with deictic referents like the anaphoric marker *ri …ri* in Kaqchikel, the emerging definite article does not necessarily extend its functions towards the expression of deictic referents.

### 4.1.4 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 from the discourse situation. In some languages, for instance in English, the definite article is able to mark such referents even without any previous mention:

(39) **The moon** is bright tonight.

English (prim. data)

Because this holds for many definite articles found in Europe, the use of definite articles has often been related to the coding of absolutely unique referents (e.g. Dryer 2013a; Dryer 2014; Hawkins 1978; Löbner 1985) as well, the uniqueness of the referent being the explanation for the presence of the definite article. In Hawkins (1978: 115), such referents are grouped with contextually unique referents as “larger situation uses” of definiteness markers and also Löbner (1985) subsumes them
under the noun type of “individual nouns”. This shows that the presence of the definite article with this type of referents is expected, if not a necessary criterion to treat a given marker as a definite article. In this section, I argue that although many languages of the world might require the use of the definite article with such referents, this is not necessarily the case and such contexts should not be taken as a necessary criterion for definite articles.

One reason to treat absolutely unique referents differently from other types of identifiable referents is that they consist of only a small (and probably) restricted set. Usually, examples are restricted to ‘sun’ and ‘moon’, but, depending on cultural and ethnographical settings, we can extend the set to earth, ocean/sea, rain, God. Also, sometimes the expression for the own tribe/people is counted in, but even though some more referents might be added to this set, it seems to be confined to a rather small number of members. For the use of the definite article with nouns expressing these referents, this means that we often deal with conventionalized uses. Since the set of referents can hardly be extended, we cannot speak of a productive use of the definite article in such contexts, the number of absolutely unique referents being restricted to a small number of nouns.

Another reason to treat the expression of absolutely unique referents differently from contextually unique referents comes from the properties of definite articles across the world’s languages. Definite articles that systematically encode anaphoric, bridging, contextually unique, and establishing referents show no apparent trend towards or against the compatibility with absolutely unique referents. Making it a criterion of the definite article would arbitrarily exclude a number of definite articles that otherwise have the same properties as the remaining definite articles. Therefore, I do not take the coding of absolutely unique referents as a criterion of definite articles, but as a function they may or may not have.

The crosslinguistic tendency of definite articles to mark absolutely unique referents or not is difficult to determine at this point, given that many grammars do not provide any information or examples for this question. Nevertheless, from the cases in my sample that do mention the use of definite articles in such contexts, no trend is evident. The following examples illustrate both types of definite articles with respect to their compatibility with absolutely unique referents. Examples (40) to (42) show definite articles in different languages that are used with absolutely unique referents.

(40) ọ ẓọngọ ịgbẹ/*ịgbẹ
2SG NEG. look sun ART:DEF sun
'Don’t look into the sun.’

Mokpe (prim. data)
(41) ʔírf-is múlmíl-ind
   moon-ART:DEF round-F
   ‘The moon is round.’
   Dime (Seyoum 2008: 44)

(42) [kui na] ia-re a-e-kiu-ge malav va suni kini
   sun.M ART:DEF.M be-hot-NF 3SGM.O-SBD-die-ANT people.PL ART:DEF.PL all
   lagí-re lo-?
   shelter-NF finish-PL
   ‘The sun was very hot, and the people all went and sheltered (in the shade).’
   Lavukaleve (Terrill 2003: 91)

In (43) to (45), we see examples from languages with definite articles which do not occur with absolutely unique referents.

(43) áxxaashe it asii-ssaa-k
   sun yet emerge-NEG-DECL
   ‘The sun hasn’t come up yet.’
   Crow (Graczyk 2007: 228)

(44) sáp-èt á:n-síʃ ha-pòn-pòn-á.
   night-LOC moon-SBJ AFF-go.out-RED-DECL
   ‘The moon comes out at night.’
   Bambassi (Ahland 2012: 445)

(45) Ken hadali min-ka kho there-n hibin
   and sun minimum-PERF not hot-SUB already
   ‘And the sun was already very hot.’
   Arawak (Pet 2011: 206)

4.1.5 Definite articles with proper nouns

Similarly to the compatibility of definite articles with absolutely unique referents, their occurrence with proper nouns (including names) is not a criterion for their classification as definite article or not.

As for definite articles, we may expect them 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. Crosslinguistically, we find examples supporting both hypotheses, with no evident trend either. In some languages, definite articles occur with names and proper nouns, while their co-occurrence is blocked in others. We can distinguish between three co-occurrence patterns of definite articles and names. The article can either be systematically required with names, its use
can be optional, and it can have a different exponent with names and proper nouns in contrast to contexts with common nouns.

Examples (46b), (47b), and (48b) illustrate the co-occurrence of the definite article with names or proper nouns in Arawak, Irish, and Tarahumara. The use of the definite article with (certain types of) proper nouns is generally required in these languages.\(^5\)

(46) a. \textit{li falhetho}  
\textsc{art:def} white.man  
‘the white man’

\hfill b. \textit{li Wim}  
\textsc{art:def} Bill  
‘Bill’

Arawak (Pet 2011: 52-53)

(47) a. \textit{an bád}  
\textsc{art:def} boat  
‘the boat’

\hfill b. \textit{an Cháisc}  
\textsc{art:def} Easter  
‘Easter’

Irish (Stenson 2008a: 35-36)

(48) a. \textit{echi torí}  
\textsc{art:def} rooster  
‘the rooster’

\hfill b. \textit{echi Antonio Loera}  
\textsc{art:def} Antonio Loera  
‘Antonio Loera’

Tarahumara (Cohen 1998: 70, 129)

In addition to definite articles that occur together with names and proper nouns, we also find languages in which anaphoric articles can be used with such referents. This is shown in (49) for the anaphoric article in Lango.

(49) \textit{òkélò-mɛ́rɛ́ tɛ̂ ringó}  
\textsc{Okelo-art:ana} 3sg.then run  
‘Then the aforementioned Okelo ran’  
Lango (Noonan 1992: 251)

In other languages, definite articles can combine with proper nouns and names, but their co-occurrence is optional, as in German (50) or Papuan Malay (51). Or, definite articles may be incompatible with proper nouns, which is what we observe in, for instance, English, Sheko (Hellenthal 2010: 142), Biak (van den Heuvel 2006: 221), or Diyari (Austin 2011: 43).

(50) (\textit{der}) Thomas  
\textsc{art:def.m} Thomas  
‘Thomas’

German (prim. data)

(51) Rahab (de)
Rahab ART:DEF
‘Rahab’

Papuan Malay (Kluge 2017: 355)

In other languages, the definite (or specific or referential) article has different exponents for its use with common and proper nouns. This is often found in Austronesian languages. Example (52) illustrates this for the referential article in Rapa Nui: the article exponent occurring with common nouns that are used in referential contexts is te, while a is used together with proper nouns.

(52) he oho a Hotu ki te hare
ART:REF PROP Hotu to ART:REF house
‘Hotu went home.’

Rapa Nui (Kieviet 2017: 102)

4.2 Anaphoric articles

Anaphoric articles were defined in section 3.3.1 as articles that systematically occur definite anaphoric referents, but not with deictic or contextually unique referents. In this section, I first present an example of anaphoric articles from Limbum (Bantu, Cameroon). The second part of this section provides examples of anaphoric markers from other languages.

4.2.1 Limbum

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:

Table 4.3: Limbum demonstratives

<table>
<thead>
<tr>
<th>Closeness</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>šá (čá)⁶</td>
</tr>
<tr>
<td>H</td>
<td>ānā</td>
</tr>
<tr>
<td>S &amp; H</td>
<td>šá (čá)</td>
</tr>
</tbody>
</table>

In addition to these spatial demonstratives mentioned in Fransen (1995), Limbum has another marker fɔ̄ which qualifies as an anaphoric article: it 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 fɔ̄ is obligatory in anaphoric contexts is illustrated in examples (53), (54), and (55). In example (53), we see the anaphoric referent expressed by mdzǐp ‘water’ in A’s second

⁶The versions of the demonstratives given in brackets correspond to the forms used by my informant.
utterance. It requires the presence of the anaphoric article \( f \). In addition, the distal demonstrative \( \dot{\text{a}}n\dot{\text{a}} \) is not felicitous in this context.

\[(53)\]

A: \( \text{ndn} \dot{\text{o}} \ \dot{\text{y}} \dot{\text{u}} \dot{\text{j}} \ \dot{\text{m}} \dot{\text{e}}. \) neck is dry 1SG

‘I am thirsty.’

B: \( \dot{\text{c}} \dot{\text{r}} \ \text{mdzip} \ \dot{\text{m}} \dot{\text{o}}? \ \dot{\text{a}}? \)

you have water some Q

‘Do you have some water?’

A: \( \dot{\text{v}} \dot{\text{u}} \ \dot{\text{m}} \dot{\text{u}} \ \dot{\text{c}} \dot{\text{r}}? \ \dot{\text{m}} \dot{\text{c}} \ \dot{\text{y}} \dot{\text{e}} \ [\text{mdzip} \ '(f)\] \ k\dot{\text{a}}? \)

come DIM small I PROG see water ART:ANA NEG

‘Can you come? I do not see the water.’

\[A': \dot{\text{v}} \dot{\text{u}} \ \dot{\text{m}} \dot{\text{u}} \ \dot{\text{c}} \dot{\text{r}}? \ \dot{\text{m}} \dot{\text{c}} \ \dot{\text{y}} \dot{\text{e}} \ [\text{mdzip} \ '(\dot{\text{a}}n\dot{\text{a}})\] \ k\dot{\text{a}}? \)

come DIM small I PROG see water DEM NEG

Limbum (prim. data)

Two other examples illustrating the anaphoric use of \( f \) and at the same time the infelicitous use of the distal spatial demonstrative \( \dot{\text{a}}n\dot{\text{a}} \) in such contexts are shown in examples (54) and (55): Again, we see in (54) that the referent of \( \text{ndzíŋwɛ̀} \ ‘woman’ \) uttered by B, referring back to the same referent uttered by A has to be marked by \( f \), while the demonstrative \( \dot{\text{a}}n\dot{\text{a}} \) cannot be used to mark the referent of \( \text{ndzíŋwɛ̀} \ ‘woman’ \) as anaphoric. The same is shown for the anaphoric referent of \( \text{mú} \ ‘child’ \) in (55).

\[(54)\]

A: \( \text{mú} \ \dot{\text{m}} \dot{\text{f}} \dot{\text{e}}? \ \dot{\text{m}} \dot{\text{e}} \ \dot{\text{m}} \dot{\text{b}} \dot{\text{a}}? \), \( \dot{\text{m}} \dot{\text{e}} \ \dot{\text{m}} \dot{\text{u}} \ \dot{\text{y}} \dot{\text{e}} \ [\text{ndzíŋwɛ̀} \ *(f)\] \ \dot{\text{b}} \dot{\text{o}} \ \dot{\text{b}} \dot{\text{a}} \ \dot{\text{a}} \ \dot{\text{t}} \dot{\text{j}} \dot{\text{u}} \dot{\text{t}} \dot{\text{e}} \ \dot{\text{m}} \dot{\text{i}} \) when time I PST2 arrive I PST2 see woman and children two 3PL sit in kitchen

‘When I came home, I found a woman with two children sitting in my kitchen.’

B: \( \dot{\text{w}} \dot{\text{e}} \ \text{mú} \ \dot{\text{r}} \dot{\text{i}} \ \ [\text{ndzíŋwɛ̀} \ *(f)\] \ \dot{\text{a}}? \)

2SG PST2 know woman ART:ANA Q

‘Did you know the woman?’

\[B': \dot{\text{w}} \dot{\text{e}} \ \text{mú} \ \dot{\text{r}} \dot{\text{i}} \ \ [\text{ndzíŋwɛ̀} \ *(\dot{\text{a}}n\dot{\text{a}})\] \ \dot{\text{a}}? \)

2SG PST2 know woman DEM Q

Limbum (prim. data)

\[(55)\]

a. \( \dot{\text{àmb}} \dot{\text{b}} \ \dot{\text{d}} \dot{\text{j}} \dot{\text{o}}? \ \dot{\text{à}} \ \dot{\text{m}} \ \dot{\text{c}} \ \dot{\text{r}} \ \dot{\text{o}} \ \dot{\text{b}} \dot{\text{z}} \dot{\text{h}} \ \dot{\text{t}} \dot{\text{a}} \dot{\text{t}} \dot{\text{r}} \ \dot{\text{ê}} \ \dot{\text{y}} \ \dot{\text{o}} \ \text{mú} \ \dot{\text{m}} \dot{\text{o}}? \)

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.’

b. \( *[\text{mú} \ *(f)] \ \dot{\text{a}} \ \dot{\text{k}} \dot{\text{ê}}? \ \dot{\text{à}} \ \dot{\text{c}} \ \dot{\text{r}} \ \dot{\text{w}} \dot{\text{ê}} \)

child ART:ANA 3SG start INF PROG cry

‘The child started to cry.’
Although the marker \( f_3 \) in Limbum is systematically used to mark anaphoric referents, it is not obligatory in all anaphoric contexts. Example (56) 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 this alternative generic reading, but it also seems to involve a stylistic contrast. Since such cases are also attested in other languages with e.g. definite articles (in fact, the English translation of (56) shows the same effect for definite article in English), we can still assume that \( f_3 \) is an anaphoric article, allowing for exceptions.

(56) \( \eta k\text{ār} \ a bitī \ r\text{k}ɔ̀mni [r\text{k}ɔ̀mni (f_3)/ \ 'anā] a bitī \ rb\text{ān}i. \) 'Friendship became love. (The) love became hatred.' Limbum (prim. data)

In addition, we find some anaphoric contexts in which the anaphoric article \( f_3 \) and the demonstrative \( \text{ānā} \) can both be used and are competing against each other. Example (57) below shows that in certain cases, both the anaphoric article \( f_3 \) as well as the spatial deixtic \( \text{ānā} \) can be used:

(57) \( \eta wɛ̀ \ mɔ̀ \ a \ c̣i \ c̣u \ m\text{i} \ Nkambe. \eta wɛ̀ \ \text{ānā}/f_3 \ a \ m \ tʃur \ bō \ sàmbā ... \) 'There used to live a man in Nkambe. The man had 7 children ...’ Limbum (prim. data)

Such a situation is similar to other languages with a definite article that “competes” in certain contexts with a stronger demonstrative marker that may be used for reasons of emphasis or a number of other interacting factors such as discourse prominence of the reference, distance between the antecedent and the anaphoric referent. The same can be assumed to hold for the distribution between a spatial demonstrative and an anaphoric article in certain contexts. A more detailed analysis of the factors that condition the choice between those two markers in certain contexts would go beyond the purposes of the present study. The fact that we find this kind of variation does however not mean that Limbum \( f_3 \) is not an anaphoric article; it corresponds to the variation and competition that we find between other types of articles and demonstratives (and pronouns in general).

In addition to the previous examples showing that \( f_3 \) is systematically used to mark anaphoric referents, it can be shown that the use of \( f_3 \) is not felicitous with other types of definite referents. Example (58) shows an instance of bridging. The bridging referent, expressed by \( \eta wɛ̀ rs\text{ān} \) ‘author’,

\[c. \] \[\text{...[mù ?ānà] à kɛ̄ʔ á c̣i wānì} \]
\[\text{child DEM 3SG start INF PROG cry} \]

Limbum (prim. data)
is identifiable due to its unambiguous link to the referent of ƞwāʔ 'book'. Nevertheless, ƞ3 cannot be used to mark the referent of ƞwè rsänj 'author' as definite:

(58) Tantōh ā mū yū ƞwāʔ m3; [ƞwè rsänj (*ƞ3)] ā ƞwè-mbɔm. 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 (prim. data)

The same effect can be observed for contextually unique definite referents. The two examples (59) and (60) below illustrate that the anaphoric article ƞ3 is not felicitous with this type of referent:

(59) [ntāh (*ƞ3)] yō fē ā làʔ nā? market.cl9 ART:ANA be.cl9 where in village this? ‘Where is the market in this village?’ Limbum (prim. data)

(60) [tō ntāʔ (*ƞ3)] tvīr mfèʔ āmbò bbir kāʔ. head chair:AM ART:ANA have time about wars NEG ‘The President does not care (does not have time for) about the strikes.’ Limbum (prim. data)

Neither are recognitional referents marked by the anaphoric article ƞ. Instead, Limbum uses the distal demonstrative ǂa, as is shown in examples (61) and (62).

(61) Context: Speaker and hearer were supposed to talk to a certain woman.

mē gī ā cī dēʔ wēr [ndįŋwē ǂa /ƞ3]
I pst1 1sg prog talk with woman DEM:DIST ART:ANA ‘I just talked to that woman.’ Limbum (prim. data)

(62) wāʔ ā kūcī [mbvī ǂa] āʔ? POL you remember goat DEM:DIST Q ‘Do you remember that goat (we used to have)?’ Limbum (prim. data)

Concerning establishing referents, the anaphoric article is not necessarily used but it can occur with this type of referents. Example (63) shows ƞ3 with an establishing referent:

(63) mē bī kir ēi kōnī [ŋwēh ƞ3 zhi mē mu suŋ we āmbò ye nā] ntίnį
I fut again prog meet person ART:ANA comp I pst2 tell 2sg about 3sg cd today ‘I am meeting [that person that I told you about] again later today’ Limbum (prim. data)

To conclude this section about ƞ3 in Limbum, we saw that the marker is systematically used in anaphoric contexts but not in other types of definite contexts; importantly, it cannot mark con-
4.2.2 Akan

Another anaphoric article can be found in Akan (Kwa, Ghana). The anaphoric article no\(^7\) was addressed in different previous works and is usually referred to as a definite article/marker/determiner (Amfo 2010, Arkoh 2011, Arkoh & Matthewson 2013, Korsah 2017). I argue in this section that it should rather be treated as an anaphoric article, as it is systematically used in anaphoric but not in other definite contexts.

Example (64) shows the article no in an anaphoric context: after the referent of èkùtú ‘orange’ is introduced, it has to be mentioned together with the article in (18b). The same can be seen in example (65), where ŋbaa ‘the woman’ is mentioned anaphorically together with no.

(64)  a. Mʊ̀tɔ́-ɔ̀ èkùtú bi 1sg-buy-pst orange art:exspec ‘I bought an orange.’
   b. [Èkùtú nʊ́] yè dèw dë̀. orange art:ana be nice so ‘The orange is/was so nice.’

(65)  A: me-kɔ-duru-uf no, me-hu-u ŋbaa bi ne mmɔfra mmienu 1sg-go-arrive-pst cd 1sg-see-pst woman art:exspec with children two bi se wɔ-te mukaasese hɔ. art:exspec comp 3pl-sit kitchen — there ‘When I arrived, I saw a woman with two children sitting in the kitchen.’
   B: (Aso) wo-nim ŋbaa no (anaa)? 2sg-know woman art:ana q ‘Did you know the woman?’

Arkoh & Matthewson (2013) call the marker no a “strong definite article”\(^7\) and argue for familiarity as a condition for its application. 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 “real-life conversations conducted between the first author and consultant S.” to show this; one such example is given in (66):

\(^7\)I use the spelling of no here following the Standard orthography of Akan. In some examples from cited literature, the anaphoric article is spelled as nʊ́ which represents the IPA transcription of the vowel and the high tone. Other than that those two forms are equivalent.
Neither does the anaphoric article occur with a (spatial) deictic referent if the latter is mentioned for the first time in the discourse situation. Example (67) from Arkoh & Matthewson (2013) shows that even when the referent of bòdm 'dog' is present in the discourse situation, reference to it using the anaphoric article no is not felicitous. It results in B’s question regarding the identity of the dog referred to by A.

(67)  (Context: Out of the blue; no deictic gesture; the dog is in view at the time of utterance.)

A:  Kwèsí bò̀̀ [bòdm nû].
    Kwesi beat-pst dog ART:ANA
    ‘Kwesi beat the dog.’
B:  Êbén bòdm á?
    which dog Q PST-NEG-say dog every self case EXPL
    ‘Which dog?’

Another relevant definite referent type is a contextually unique referent that has an unambiguous link to another referent of the discourse situation and allows its identifiability at the first mention. Example (68) shows that the anaphoric article no in Akan is not used in such context: the contextually unique referent of ġuá ‘market’ is expressed by a bare noun without the article.

(68)  Mù-rô-kò ġuá mû.
     1SG.-PROG-go market in
     ‘I am going to the market.’

Similarly to contextually unique referents, bridging referents are identifiable and thus definite at their first mention because they are unambiguously linked to another, previously mentioned referent. Examples (69) and (70) illustrate this referent type; as expected, they show that the anaphoric article is not used in bridging contexts. Instead, it is the use of the possessive marker ní that expresses the link between the bridging referents kín ‘neck’ and nkyensidan ‘roof’ in (69a) and (70a), respectively:
(69) a. Ìgùán nó sò árá mà nyímpá ànán nà wó-dzi-i [nì 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-dzi-i [kón ní].
   sheep ART:ANA big just COMP person four FOC 3PL-eat-PST neck ART:ANA
   Akan (Arkoh & Matthewson 2013: 14)

(70) a. Ye-hu-u dan dadaw bi wɔ ekurasi hɔ [nì nkyɛnsidàn]
   1PL-see-PST building old ART:EXSPEC at village there POSS roof
   worn.out
   ‘We saw an old building in the village; the roof was worn out.’

b. #Ye-hu-u dan dadaw bi wɔ ekurasi hɔ [nkyɛnsidàn ní]
   1PL-see-PST building old ART:EXSPEC at village there roof ART:ANA
   ehodwow.
   worn.out
   Akan (Arkoh 2011: 80)

Given that the previous examples showed in detail that no in Akan systematically occurs with anaphoric referents but is not used with other types of definite referents, it comes without surprise that it can not be used to mark absolutely unique referents either. Examples (71) and (72) show this for the referents ‘moon’ and ‘pope’.

(71) me-n-hu [ɔsram (*no)] anumere yi
   1sg-NEG-see moon ART:ANA evening this
   ‘I will not see the moon this evening.’ Akan (prim. data)

(72) Kwámì nyá-à kràtàá fí-ì Ègyá krónkrón póp hɔ.
   Kwame get-PST letter from-PST father holy pope there
   ‘Kwame got a letter from the holy father Pope.’ Akan (Arkoh & Matthewson 2013: 11)

Establishing referents are referents that are marked as identifiable but are only made identifiable to the hearer by the immediately following discourse segment. This often involves a syntactic construction of relative clauses, and in Akan, the anaphoric article no can be used with definite referents of this type:

   1.FUT-show you DEM pictures ART:ANA rel they-come.from Egypt cd
   ‘I will show you the pictures from Egypt.’ Akan (Arkoh & Matthewson 2013: 8)
In her heart, Amma still loves her village. Especially the people, the good air, the fruits that are there. Akan (prim. data)

In addition to the use of the anaphoric article *no* together with nouns to express an establishing referent, *no* also appears as a clausal determiner (glossed *cd*) in examples (73) and (74). I will not discuss the use of *no* as a clausal determiner here; suffice it to say that it is likely that this use is an extension of the anaphoric article. This development is especially plausible assuming that establishing referents play a crucial role for the development of definite articles (Himmelmann 1997: 61-82). Since those referents typically occur in a construction involving a following restrictive relative clause in which the referent is made identifiable, it is not surprising that we find a marker in such contexts that is related to the anaphoric article.

In general, the use of the anaphoric article with establishing referents points towards a functional extension of the anaphoric article towards contexts that do not necessarily involve identifiability based on previous mention. While I treat the article *no* in Akan as an anaphoric article here, we might as well deal with an emerging definite article. Further supporting evidence for the functional extension of the anaphoric article in Akan comes from other bridging contexts in which the anaphoric article can be used; its distribution in bridging contexts is thus not categorically excluded. On the one hand, this could simply be due to the fact that bridging contexts involve semantic relations of various kinds that affect the distribution of certain referential markers in a different way: In some contexts, a possessive marker may be preferred and may block the use of an article, while in other contexts, a possessive marker may not be available and the article may be used. On the other hand, it could indeed show that we find variation with regard to the use of *no* because its function is extending from an anaphoric towards a definite article.

Examples (75) and (76) show the use of the anaphoric articles together with bridging referents. In (75), we again see the bridging relation between *a book* and *an author*; the latter is marked by the anaphoric article in (75b), while (75c) shows an alternative strategy with a possessive marker. In the latter case, the anaphoric article cannot be used. In example (76), we see bridging between the event of *arriving* that is unambiguously linked to the *bus* as medium of transport (at least in Ghanaian general knowledge). Again, the anaphoric article has to be used with the noun *baase* in this bridging context.

---

8 For more details on the clausal determiner in Akan and related languages, see Korsah (2017), Lefebvre (1992).
(75) a. Kofi tɔ-ɔ krataa bi ...  
Kofi buy-pst book ART:EXSPEC ...  
‘Kofi bought a book …’

b. …[otwɛfɔɔ no] fi Kumase.  
…writer ART:ANA come.from Kumasi  
‘…The author is from Kumasi.’

c. …[ne-twerɛfɔɔ (*no)] fi Kumase  
…poss:3sg-writer ART:ANA come.from Kumasi  
‘…Its author comes from Kumasi.’

(76) seesia na wɔ-fri Nkran be-du-ue. [baase no] ye-ɛ lati paa.  
now foc 3sg-leave Accra fut-reach-pst bus ART:ANA be-pst late very  
‘They just arrived from Accra. The bus was very late.’

4.2.3 Komnzo

Komnzo (Morehead-Wasur, Papua New Guinea) has the anaphoric article *ane*; its use with anaphoric referents is shown in examples (77) and (78):

(77) a. wämne ...yf füni yé firra=n bā ykogr.  
tree ...name füni be:s:3SG.M.NPST,IPFV firra=LOC DEM:MED  
‘The name of the tree is füni. It stands there in Firra.’

b. nā kayé fthé boba gnyako nima kwa  
ART:PREs yesterday when DEM:MED:ALL go:s:3SG:IMP like:DEM  
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)

(78) a. bthan kabe fthé fenz yona-si bänemr zrethkāfth  
magic man when body.liquid drink-NMLZ RECOG.PURP start..IRR.PFV.S:3PL  
mātrak-si=t.  
take.out-NMLZ=PURP  
‘When the sorcerers drink the body fluids, they start by bringing out this one.’

b. …fthé fo[ emph krefar] bobo fokam zn=fo  
when EMPH set.off.IRR.PFV.S:3SG ART:ANA magic man MED:ALL grave place=LOC  
fokam mnnz=fo sikwankwan=me zbār 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)
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.”

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 (79).

(79) fintäh ane z=yé yem=aneme dagon.
    fintäh ART:ANA DEM:PROX=be.3sg.NPST cassowary=POSS:NSG food
    ‘This fintäh (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 (80), the referent of ṇars ‘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. In (81), the preceding context sets the scene of rubbish being put into the house of the woman, which she then cleans. In the contexts of this given house, the referent of dödö ‘broom’ is contextually unique by default, assuming the household has a single broom, and the referent is thus identifiable as well, being mentioned for the first time. Again, we see that the anaphoric article ane is not used in this context.

(80) kabe matak erä nima z bramöwá kwafarkwrth nima man nothing be.NPST:IPFV.3PL like.this already all set.off.PST:IPFV.3PL like.this erä ṇars=fo ... be.NPST:IPFV.3PL river=ALL ...
    ‘Nobody is here. All the people left this way to the river ...’ Komnzo (Döhler 2018: 365)

(81) dödö thfefa ane zurenwrmə 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. However, Komnzo has a separate recognitional marker baf that is used only pronominally (Döhler 2018: 112). For establishing referents, on the other hand, we find ample evidence in Döhler (2018) that ane is used to mark them as identifiable. Examples (82) and (83) illustrate the use of the anaphoric article ane in establishing contexts, occurring with referents that have not been mentioned in the previous discourse segments.
(82) nagayé=aneme znsä=n zwäfonz  
children=poss.nsg work=loc be.caught.by.nightfall.recpst:pfv.1sg art:ana

gathagatha=me k-kauna mane egathikvroth].
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)

(83) waisamen ...waisamen [ane kar yf rä mobo
waisamen ...waisamen art:ana place name be.npst:ipfv.3sg.f where.all
zwamnzrm] ...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)

4.3 Recognitional articles

Across the world’s languages, we find another type of markers that qualifies as an article. These
markers are usually discussed as “recognitional” demonstrative or “recognitional function” of
demonstratives (e.g. Himmelmann 1997: 61-72, Diessel 1999: 105-109). What is meant by recog-
nitional function is the identifiability of a referent by both the speaker and and hearer based on
previous (common) knowledge or mutual experience in the in past, which was defined as a recog-
nitional referent in section 3.2.1.

In many languages of the world demonstratives are used to mark recognitional referents.
Demonstratives that mark recognitional referents are not treated as recognitional articles. I only
consider markers that are exclusively used to encode recognitional referents as recognitional ar-
ticles. The motivation for treating them as articles rather than another type of demonstratives
follows the motivation for treating anaphoric markers as articles: marking recognitional refer-
ents is an abstract referential function, and dedicated markers are rather rare crosslinguistically.

Oksapmin (Oksapmin, Papua New Guinea) is an example of a language with a recognitional
article. 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 (137) and (85). Loughnane (2009: 124) notes for example (137): “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”:
gin i ml-sa jaxe 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)

The context of example (85) 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):

(85) [niu jia max] baten 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)

Gooniyandi (Bunuban, Australia) features a recognitional article as well, occurring with definite referents that are identifiable by knowledge shared between the speaker and the hearer. According to McGregor (1990: 146), the marker is used for referents from shared previous experiences or conversations, as is shown in example (86).

(86) ngooddo-ngga [ginharndi goornboo] wardgilayi-ngangi mooooloodja-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)

Yankuntjatjara (Desert Nyungic, Australia) is another language with a recognitional article. Although treated as an anaphoric marker in Goddard (1985: 54), the description of its use corresponds to the coding of recognitional referents. An example is shown in (87).

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)

(87) munu nyaku-payi “waru kampa-nyi [kapi panyayngka-manti-r]
ADD see-CHAR fire.NOM burn-PRES water ART:RECOG-LOC-PROB-EXCIT
‘And they would see “There’s a fire burning, at that water(hole), you know the water(hole), most likely, by gee.”’

Yankuntjatjara (Goddard 1985: 54)
Himmelmann (1997: 69) notes that the marker *nhenge* in Arrernte (Arandic-Thura-Yura, Australia) is glossed as ‘remember’ in the grammar (Wilkins 1989: 121) and used in recognitional contexts. In Wilkins (1989: 121), its use is described 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 following two examples to illustrate the two uses: example (88) shows *nhenge* marking an anaphoric referent (not adnominally in this example), while example (89) demonstrates that *nhenge* can also encode a recognitional referent:

(88)  ...kem-irre-ke  thipe kngerrepene 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)

(89)  [inspector *nhenge*]  map-le school nhenhe-rlke inspect-emimle-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)

Example (88) shows that Arrernte *nhenge* does not only mark recognitional definites, but would be better described as an anaphoric marker whose use can extend to recognitional contexts. Therefore, I do not treat *nhenge* as recognitional article here.

This section presented a number of recognitional articles. Their occurrence seems particularly favored in Australia and Papunesia. Since little work has been done on this referential function or on the markers that encode recognitional referents, future research is required in order to give a more detailed picture of recognitional articles.

### 4.4 Summary

This chapter presented different examples of articles in the definite domain. I argued that the occurrence of definite articles with absolutely unique and deictic definites may be a common but not necessary property. The use of definite articles together with spatial deictic referents was
shown to be dependent on the properties of the source element, a spatial demonstrative in most cases. This could explain that many definite articles can be used to encode deictic referents: spatial demonstratives are the most frequent source for definite articles. In addition to definite articles, I showed that two other types of articles can be distinguished in the definite domain: anaphoric and recognitional articles. Anaphoric articles only encode anaphoric referents; recognitional articles mark referents that are definite based on previous experience shared between the speaker and the hearer.
Chapter 5

Articles in the indefinite domain

This chapter provides an overview of different article types from the indefinite domain by discussing examples of exclusive-specific (section 5.1), nonspecific (section 5.2), and indefinite articles (section 5.3). 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 in North America. In section 5.2.4, I show that some nonspecific articles can be traced back to verbal irrealis markers. The steps of the development from a verbal irrealis marker to a nominal nonspecific marker include: occurring on nouns in the predicate position to encode irrealis-related meanings, occurring on nouns in argument position to encode irrealis-related meanings, and occurring on nouns marking their referents as nonspecific.

The second observation concerns a subtype of indefinite articles that I call presentational articles. We find many emerging indefinite articles that are restricted to discourse-prominent referents. These are not necessarily restricted to specific referents, but can also occur with nonspecific referents. This is important insofar as the grammaticalization path traditionally that is traditionally presented for indefinite articles relate the extension from highly discourse-prominent referents to less prominent referents to that of specific to nonspecific referents. By discussing the use of such presentational articles in section 5.3.4, I show that the two developments are in fact correlated but independent from each other.

5.1 Exclusive-specific articles

This section deals with exclusive-specific articles. As defined in section 3.3.2, 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 nonidentifiable by the speaker, as in I was attacked by a penguin.
yesterday. Following e.g. Giannakidou (1997, 2017), I take contexts that refer to single past events as episodic contexts and the nonidentifiable referent that is involved as specific.

5.1.1 Q’anjobal

Q’anjobal has a marker jun that, as I will argue in this section, is an exclusive-specific article. In the literature, it is usually labelled an indefinite article, e.g. in Mateo Toledo (2017) and Raymundo Gonzáles et al. (2000). However, none of the examples discussed in those two sources include nonspecific referents. In addition, we find dialectal variation in Q’anjobal with respect to the exponents of the plural form of the article (Raymundo Gonzáles et al. 2000: 153), as is shown in Table 5.1.

<table>
<thead>
<tr>
<th>region</th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soloma &amp; Ixcoy</td>
<td>jun</td>
<td>jan</td>
</tr>
<tr>
<td>Santa Eulalia &amp; Barillas</td>
<td>jun</td>
<td>juntzan</td>
</tr>
<tr>
<td>San Rafael &amp; Acatán</td>
<td>jun</td>
<td>wan</td>
</tr>
</tbody>
</table>

Table 5.1: Plural exponents of the exclusive-specific article in Q’anjobal

The examples presented in this section are based on the variety spoken in Santa Eulalia (Huehuetenango, Guatemala). Future research is required to show whether the distinction between a specific-exclusive and a nonspecific article in Q’anjobal is only present in this variety or also available in others. I argue in this section that jun is not an indefinite article, since it does not occur with nonspecific referents. The latter are marked by the nonspecific article junq (see section 5.2.1).

Examples (1) and (2) show the marker jun in specific indefinite contexts, occurring together with the expressions lugar ‘place’ and -tut ‘house’. The referents of ‘place’ and ‘house’ are specific because the expressions uttered are linked to particular referents from their kind sets, but they are not identifiable by the hearer.

(1) wojtaq jun lugar baytal chije’ ko-kuywi
    KNOW.1SG ART:EXSPEC place where can 1DU-study
    ‘I know a (certain) place where we can study.’ Q’anjobal (prim. data)

(2) cham anima may kon tojlaneni ay jun miman ya-tut
    CL:male person PST COMP pay.PST:S:3SG.O:1PL EXIST ART:EXSPEC big 3SG-house
    ‘The man who paid us has a big house.’ Q’anjobal (prim. data)

The specific article jun in Q’anjobal has the plural form juntzan, which is illustrated in (3) and (4). In (3), its presence marks the referent of watz’il-anima ‘friend’ as specific. In the existential
construction in (4), the article juntzan is also used to mark the referent of qajaja ‘mosquito’ as specific. In this case, the referent is marked as particular but not identifiable simply because its full identification is not relevant in the given discourse situation.

(3) ay [juntzan heb’ hin-watz’il-anima] max hin ul yulanej yet exist ART:EXSPEC.PL PL 1SG-friend-person PST me come.PST.3PL visit.PST.3PL TEMP ma’i aq’ab’alil today night ‘Some friends (of mine) visited me last night.’ Q’anjobal (prim. data)

(4) yiban te mexha ay [juntzan qajaja] above CL:wood table exist ART:EXSPEC.PL mosquito ‘There are (some) mosquitoes (flying) above the table.’ Q’anjobal (prim. data)

In order to be a specific article, jun must be incompatible with nonspecific referents. Examples (5) and (6) show jun in indefinite specific and nonspecific contexts. The referent of yatut thioxh ‘church’ in (5) is specific, as it is a particular referent of its kind that is linked to the expression used. In this case, only the article jun is felicitous. In example (6), on the other hand, the referent of yatut thioxh ‘church’ is nonspecific because it falls under the scope of the question, so that any referent from the kind set could be linked to the expression. In such a context, the specific article jun cannot be used, and the nonspecific article junoq occurs instead.

(5) 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’anjobal (prim. data)

(6) 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?’ Q’anjobal (prim. data)

It is very plausible that the nonspecific article junoq is diachronically related to the specific article jun. Therefore, one could argue that we are dealing with only one marker jun used in indefinite specific and nonspecific contexts, the latter requiring an additional marker -oq. Even though the markers are related diachronically and even though the nonspecific marker can formally be decomposed into the specific article and an additional marker -oq, I treat them as separate articles. I do so because, synchronically, we see two formally distinct markers with two distinct referential functions. Neither the diachronic nor the synchronic formal relation between these two markers is relevant to the question of whether they correspond to a single or two separate articles as long as we deal with two formally and functionally distinct markers.

In examples (7) and (8), we see a similar pair of specific and nonspecific referents, marked by the specific and nonspecific articles, respectively. In (7), the speaker can identify the referent of
'tuktuk', as is indicated by the continuation in the second sentence; the referent is therefore particular and specific. As expected, only \textit{jun} can be used. In (8) on the other hand, no particular referent of the kind set of 'tuktuk' is linked to the expression \textit{tuktuk}. Again, we see that \textit{jun} is not felicitous, and that \textit{junoq} is used instead.

(7) tzeb'ach yul [\textit{jun}/*junoq tuktuk] mayal wawrtej naq t'umon ch'en come.imp in \textit{ART:EXSPEC ART:NSPEC tuktuk} already called cl driver
   'Let's take a (certain) tuktuk. I already called the driver.' Q'anjobal (prim. data)

(8) asi' yul [junoq/*\textit{jun} tuktuk] go.imp in \textit{ART:NSPEC ART:EXSPEC tuktuk}
   'Let's take a (any) tuktuk.' Q'anjobal (prim. data)

The following examples illustrate the absence of the specific article with definite referents. In (9), the referent of \textit{itaj} 'herb' is deictic, and hence definite. The noun \textit{itaj} occurs together with a demonstrative and a classifier, the article \textit{jun} being absent. Example (10) illustrates an anaphoric referent and its incompatibility with \textit{jun}. Neither does the article \textit{jun} occur with an establishing referent, as can be seen in (11), where the referent is established as identifiable by the absence of the article.

(9) \textbf{Context: S is pointing at her food.}

manxa watz'ilq chiji an \textit{itaj} ti'
EMPH delicious eat cl:herb vegetables dem:prox
'These vegetables are delicious.' Q'anjobal (prim. data)

(10) A: ayit max inabniok pai in-na maxwuil \textit{jun} ix
   when pst arrive.pst.1sg.at 1sg-house see.pst.s:1sg.o:3pl \textit{ART:EXSPEC \textit{woman}}
yetuq kawan unin tolochotanayeb yul kusina
   with two child be.seated in kitchen
   'When I arrived at home, I found a woman with two children sitting in the kitchen.'

B: hojtaqmi pax [ix ix tu']?
know.2sg aff cl:female woman dem:dist
'Did you know that woman?'

(11) a [xal ix] maktet ichaj'el yuj Lwin yet ewo tolto max awji
   foc cl:polit \textit{woman} who was.met by Pedro temp yesterday just pst call.3sg
   'The woman Pedro met yesterday called not long ago.' Q'anjobal (prim. data)

Thus, this section showed that \textit{jun} cannot occur in either definite or nonspecific contexts but has to occur with specific referents, which makes it an exclusive-specific article.
5.1.2 Akan

Akan (Kwa, Ghana) also has an exclusive-specific article, *bi*, that was discussed as a marker of specificity in the previous literature, for instance in Arkoh (2011). This section shows how *bi* in Akan is systematically used in specific indefinite contexts but does not occur with definite or nonspecific referents. In (12), we see *bi* in the typical specific contexts, introducing a new referent to the discourse:

(12) da *bi* [ɔbea *bi*] ne ne ba ɔsɔodenfo *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 (14), (15), and (13) show that the exclusive-specific article is used in other types of contexts involving specific referents as well. In these examples, in contrast to (12), the specific referent has less discourse prominence. The article *bi* occurs together with the expressions *akyɛde* ‘gift’, *ahwahwadeɛ* ‘miracle’, and *mayɔnkofoɔ* ‘my friends’, which are all linked to a particular referent that is not identifiable by the hearer. In (13), the referent becomes identifiable by the hearer at a later point in time; the context for (14) excludes the hearer from knowing about the ‘gift’, and the identifiability of the specific referent in (15) is simply not relevant in the given situation.

(13) 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 (prim. data)

(14) me-wo [akyɛde *bi*] ma wo
1sg-have gift ART:EXSPEC give 2sg
‘I have a gift for you.’

Akan (prim. data)

(15) [m-ayɔnkofoɔ *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 (prim. data)

Examples (16) and (17) show that *bi* cannot be used with nonspecific referents. In (16), the referent of *yere* ‘wife’ is nonspecific because it is within the scope of the irrealis reading of the predicate; (17) shows the same in the context of a question.

Examples (16) and (17) show that *bi* cannot be used with nonspecific referents. In (16), the referent of *yere* ‘wife’ is nonspecific because it is within the scope of the irrealis reading of the predicate; (17) shows the same in the context of a question.
In order for *bi* to be an exclusive-specific article, it also has to be incompatible with definite referents. That Akan uses a different, anaphoric article in some types of definite contexts was already shown in section 4.2.2. Example (18) is repeated below to illustrate that *bi* does not occur in definite contexts and hence is another examples of an exclusive-specific article. The referent of *èkùtú* ‘orange’ is only marked by *bi* when it is mentioned for the first time, when referred back to its referent, it is expressed by the noun accompanied by the anaphoric article *no*.

(18) a. Mʊ̀tɔ́-ɔ̀ [èkùtú *bi]*
   1sg-buy-pst orange *ART:EXSPEC*
   ‘I bought an orange.’

b. [Èkùtú *nô]* yè dèw dèè.
   orange *ART:ANA* be nice so
   ‘The orange is/was so nice.’

5.1.3 Palula

Palula (Dardic, Pakistan) is another language with an exclusive-specific article. Examples (19), to (22) show that the article *áa/ák*\(^1\) marks the referent as specific, i.e. as particular but not identifiable by the hearer.

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

(20) miír thani [ák mîiš] heensîl-u de
   Mîr *QUOT ART:EXSPEC* man stay.PFV-M.SG PST
   ‘There was a man called Mîr.’

While examples (19) and (20) illustrated the use of the exclusive-specific article in the typical specific contexts of the introduction of a new, prominent discourse referent, (21) and (22) below show that it is also used with less prominent specific discourse referents.

---
\(^1\)The different forms are used in different varieties of Palula (Liljegren 2016: 140).
(21) eesé zanɡal-í [áa bat-á] jhuli hari so kunaák
   rem forest-OBL ART:EXSPEC stone-OBL on take.away.CV DEF.NOM.M.SG child.NOM.M.SG
   bheesóol-u seat.PFV-M.SG
   ‘In that forest he took the child to a stone and seated him.’ Palula (Liljegren 2016: 283)

(22) [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 (23) to (25) below show three contexts with nonspecific referents which are all expressed as bare nouns, namely fláiṭ ‘flight’, wása ‘strength’, and muloó ‘mullah’, respectively.

(23) čúur reet-i jheez-ií fláiṭ 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)

(24) tasíi 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)

(25) muloó díí 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 (26) below shows an anaphoric referent which is expressed by a noun together with an anaphoric article in Palula. The article aa/ak cannot be used in definite contexts; we can thus conclude that it indeed is an exclusive-specific article.

(26) théeba [se ṭhaatáak-a] bi tas sanɡí khainií široó thiil-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)
5.2 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. Sections 5.2.1 and 5.2.2 discuss two examples of nonspecific articles in Q’anjobal and Tongan. In the case of Q’anjobal, section 5.1.1 showed that the language also has an exclusive-specific article, while Tongan (Austronesian, Tonga) has an additional inclusive-specific article (cf. section 6.1.2). The third example of nonspecific articles comes from Lakota (Siouan, USA). Lakota has a more complex article system consisting of a definite article, an exclusive-specific article, and the nonspecific article discussed in this chapter. Finally, section 5.2.4 discusses the development of nonspecific articles in a number of languages from North America, arguing for their development from irrealis markers of the verbal domain.

5.2.1 Q’anjobal

Q’anjobal (Mayan, Guatemala) features the marker junoq that I treat as a nonspecific articles. As was mentioned in section 5.1.1, previous work on Q’anjobal only mention the marker jun, which is regarded as an indefinite article (Mateo Toledo 2017, Raymundo Gonzáles et al. 2000). Although neither of these grammatical sketches include contexts with nonspecific referents similar to the contexts shown here, Mateo Toledo (2017) mentions the irrealis marker -oq that can be related to the nonspecific article (cf. section 5.2.4). The discussion in this section is based on my primary data collection in Santa Eulalia (Huehuetenango, Guatemala).

As was already shown in the previous sections, the nonspecific article junoq occurs in complementary distribution with the specific article. Examples (27a) in opposition to (27b), as well in (28a) and (28b) illustrate this.

(27) 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.’

(28) a. tzeb’ach yul [jun/ *junoq tuktuk] mayal wawrtej naq tz’umon ch’en
come.imp in art:exspec art:nspec tuktuk already called cl 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’anjobal (prim. data)

In (29) to (32) below, I show further examples of the use of *junoq* in other nonspecific contexts. Note that the possessive marking of the nouns does not interfere with the nonspecific referents of *ha-tz’ib’al* (‘your’ pen’ and *ha-na* (‘your’) house’.

(29) ma ay [junoq ch’en ha-tz’ib’al]?  
NEG exist ART:NSPEC CL:stone 2sg-pen  
‘Do you have a pen? (Any pen will do.)’  
literally: ‘Don’t you have a (any) pen?’  
Q’anjobal (prim. data)

(30) ay [junoq ha-na]?  
EXIST ART:NSPEC 2sg-house  
‘Do you have a house?’  
Q’anjobal (prim. data)

(31) chin q’an [junoq watz jatnemajna]  
1sg need.1sg ART:NSPEC good architect  
‘I need a good architect (any architect who is good).’  
Q’anjobal (prim. data)

(32) 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’anjobal (prim. data)

Like the specific article *jun*, the nonspecific article *junoq* can also be used to express plurality. This is illustrated in example (33).

(33) a.  aq’ [junoq/ ‘jun hin-xila]  
give.IMPER ART:NSPEC ART:EXSPEC 1sg-chair  
‘Give me a (any) chair.’  
Q’anjobal (prim. data)

b.  aq’ [juntzanoq/ *juntzan hin-dulce]  
give.IMPER ART:NSPEC.PL ART:EXSPEC.PL 1sg-candy  
‘Give me some candy. (It can be any of the candies that are in the bowl.)’  
Q’anjobal (prim. data)

Having established that in the Santa Eulalia variety of Q’anjobal, the marker *jun* systematically occurs in specific contexts while *junoq* systematically marks nonspecific referents, we can assume that they are an exclusive-specific and a nonspecific article, respectively.
5.2.2 Tongan

Tongan has a nonspecific article, *ha*, in addition to the an inclusive-specific article *(h)e*. The latter will be discussed in section 6.1.2; in this section, I present examples that show the distribution of the nonspecific article *ha*.

In previous studies, the Tongan articles were regarded as definite and as indefinite articles (cf. Churchward 1985: 23-27, Morton 1962) following the traditional division made for many European languages. Although still labelling them definite and indefinite articles, Otsuka (2000: 50) notes that the use of the articles in Tongan differs from their English counterparts. Völkel (2010), on the other hand, treats the two articles as specific and nonspecific articles. Following Völkel (2010) and based on their use and distribution, I treat the Tongan articles as inclusive-specific and nonspecific articles (cf. section 6.1.2 for a discussion of the nonspecific article). The paradigm of the articles in Tongan (Völkel 2010: 159) is given in Table 5.2 below. 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 here.

<table>
<thead>
<tr>
<th></th>
<th>neutral</th>
<th>affective</th>
</tr>
</thead>
<tbody>
<tr>
<td>specific</td>
<td><em>(h)e</em></td>
<td>si’i</td>
</tr>
<tr>
<td>nonspecific</td>
<td><em>ha</em></td>
<td>si’a</td>
</tr>
</tbody>
</table>

Examples (34), (35), and (36) illustrate the use of *ha*. In (34) and (36), the referents are marked by the nonspecific article *ha*, since there is no particular referent from the kind set that would be linked to the expression. Nonidentifiable referents within the scope of imperatives are have to be nonspecific, since it seems implausible from a communication point of view to refer to a particular referent which is not identifiable by the hearer within an imperative. Example (36) also shows that the nonspecific article is compatible with plural nouns. In (35), the context allows for the referent of *vai* ‘water’ to be specific or nonspecific. Given that it is presented as nonspecific in the translation and given that it is marked by *ha*, we can assume that the nonspecific article marks the referent of *vai* ‘water’ as nonspecific here as well.

(34) *ha’u mo* [ha afo]  
    come with ART:NSPEC fishing.line  
    ’Bring a (any) fishing line.’

Tongan (Churchward 1985: 25)

---

2The form *(h)e* of the neutral specific article only occurs after certain prepositions, otherwise the form *e* is used.
(35)  'oku ou fie ma’u [ha vai mafana]
    PRS  I  want  receive  ART:NSPEC  water  warm
    ‘Is there any warm water here?’ literally: ‘I want to receive warm water.’
    Tongan (Churchward 1985: 24)

(36)  fai [ha 'u tohi]
    do.IMPER ART:NSPEC PL letter
    ‘Write (some) letters.’
    Tongan (Churchward 1985: 24)

That *ha* is also used to mark nonspecific referents in questions is shown in (16). We also see that
*ha* cannot be used with specific referents in (38); in order to mark a referent as specific, (38) shows
that the inclusive-specific article *he* must be used instead of *ha*.

(37)  'oku 'i ai [ha maa]?
    PRS  in  there  ART:NSPEC  bread
    ‘Is there some bread?’
    Tongan (Otsuka 2000: 50)

(38)  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)

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). The paradigms of these two sets are
presented in Table 5.3 and Table 5.4, respectively.
### Table 5.3: Paradigm of Tongan adnominal possessive markers (A-set)

<table>
<thead>
<tr>
<th></th>
<th>specific</th>
<th></th>
<th>nonspecific</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>neutral</td>
<td>affective</td>
<td>neutral</td>
<td>affective</td>
</tr>
<tr>
<td>1SG</td>
<td>(he)‘eku</td>
<td>i‘eku</td>
<td>ha‘aku</td>
<td>i‘aku</td>
</tr>
<tr>
<td>2SG</td>
<td>ho‘o</td>
<td>i‘o</td>
<td>ha‘o</td>
<td>i‘ao</td>
</tr>
<tr>
<td>3SG</td>
<td>(he)‘ene</td>
<td>i‘ene</td>
<td>ha‘ane</td>
<td>i‘ane</td>
</tr>
<tr>
<td>1DU.EX</td>
<td>(he)‘ema</td>
<td>si‘ema</td>
<td>ha‘ama</td>
<td>si‘ama</td>
</tr>
<tr>
<td>1DU.IN</td>
<td>(he)‘eta</td>
<td>si‘eta</td>
<td>ha‘ata</td>
<td>si‘ata</td>
</tr>
<tr>
<td>2DU</td>
<td>ho‘oma</td>
<td>si‘omo</td>
<td>ha‘amo</td>
<td>si‘amo</td>
</tr>
<tr>
<td>3DU</td>
<td>(he)‘ena</td>
<td>si‘ena</td>
<td>ha‘ana</td>
<td>si‘ana</td>
</tr>
<tr>
<td>1PL.EX</td>
<td>(he)‘emau</td>
<td>si‘emau</td>
<td>ha‘amau</td>
<td>si‘amau</td>
</tr>
<tr>
<td>1PL.IN</td>
<td>(he)‘etau</td>
<td>si‘etau</td>
<td>ha‘atau</td>
<td>si‘atau</td>
</tr>
<tr>
<td>2PL</td>
<td>ho‘omou</td>
<td>si‘omou</td>
<td>ha‘amou</td>
<td>si‘amou</td>
</tr>
<tr>
<td>3PL</td>
<td>(he)‘enau</td>
<td>si‘enau</td>
<td>ha‘anau</td>
<td>si‘anau</td>
</tr>
<tr>
<td>IMPERS</td>
<td>(he)‘ete</td>
<td>si‘ete</td>
<td>ha‘ate</td>
<td>si‘ate</td>
</tr>
</tbody>
</table>

### Table 5.4: Paradigm of Tongan adnominal possessive markers (O-set)

<table>
<thead>
<tr>
<th></th>
<th>specific</th>
<th></th>
<th>nonspecific</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>neutral</td>
<td>affective</td>
<td>neutral</td>
<td>affective</td>
</tr>
<tr>
<td>1SG</td>
<td>hoku</td>
<td>si‘oku</td>
<td>haka</td>
<td>si‘aku</td>
</tr>
<tr>
<td>2SG</td>
<td>ho</td>
<td>si‘o</td>
<td>hao</td>
<td>si‘ao</td>
</tr>
<tr>
<td>3SG</td>
<td>hono</td>
<td>i‘ono</td>
<td>hano</td>
<td>si‘ano</td>
</tr>
<tr>
<td>1DU.EX</td>
<td>homa</td>
<td>si‘oma</td>
<td>hama</td>
<td>si‘ama</td>
</tr>
<tr>
<td>1DU.IN</td>
<td>hota</td>
<td>si‘ota</td>
<td>hata</td>
<td>si‘ata</td>
</tr>
<tr>
<td>2DU</td>
<td>homo</td>
<td>si‘omo</td>
<td>hamo</td>
<td>si‘amo</td>
</tr>
<tr>
<td>3DU</td>
<td>hona</td>
<td>si‘ona</td>
<td>hana</td>
<td>si‘ana</td>
</tr>
<tr>
<td>1PL.EX</td>
<td>homau</td>
<td>si‘omau</td>
<td>hamau</td>
<td>si‘amau</td>
</tr>
<tr>
<td>1PL.IN</td>
<td>hotau</td>
<td>si‘otau</td>
<td>hatau</td>
<td>si‘atau</td>
</tr>
<tr>
<td>2PL</td>
<td>homou</td>
<td>si‘omou</td>
<td>hamou</td>
<td>si‘amou</td>
</tr>
<tr>
<td>3PL</td>
<td>honau</td>
<td>si‘onau</td>
<td>hanau</td>
<td>si‘anau</td>
</tr>
<tr>
<td>IMPERS</td>
<td>hoto</td>
<td>si‘oto</td>
<td>hato</td>
<td>si‘ato</td>
</tr>
</tbody>
</table>

An example of the use of possessive forms from these two sets is provided in (39). In both (39a) and (39b), possessive markers from the O-set are used. In the question in (39a), we see the nonspecific possessive, while the specific possessive is featured in the answer in (39b).
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)

5.2.3 Lakota

Another example of a nonspecific article is found in Lakota (Siouan, USA), occurring in complementary distribution with the specific article in indefinite contexts. This is shown in example (40). In (40a), the specific referent of *igmu’-cat’ is encoded by the exclusive-specific article *wã*. In contrast, the marker *wãži* in (40b) signals that *igmu’* has a nonspecific referent.

(40) a. [igmu’ wã] wachi 
   cat ART:EXSPEC want.s:1SG.O:3SG 
   ‘I want a (particular) cat.’

b. [igmu’ wãži] wachi 
   cat ART:NSEPSEC want.s:1SG.O:3SG 
   ‘I want a (any) cat.’

Lakota (Van Valin 1977: 64)

Since this distribution of *wã* and *wãži* with respect to specific and nonspecific contexts was reported to be systematic (Rood & Taylor 1996, Van Valin 1977, Williamson 1984), I treat *wãži* as a nonspecific article.3 The full paradigm of the nonspecific and specific article is given in Table 5.5 below.4

Table 5.5: Lakota specific and nonspecific article paradigm

<table>
<thead>
<tr>
<th></th>
<th>specific</th>
<th>nonspecific</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td><em>wã</em></td>
<td><em>wãži</em></td>
</tr>
<tr>
<td>PL</td>
<td><em>eya’</em></td>
<td><em>etã</em></td>
</tr>
</tbody>
</table>

Examples (41) to (44) are further illustrations of the distribution of the two articles in Lakota. In (41), the referent of *makno’xløva- cave’ is specific and therefore marked with the specific article. In (42) and (43), the context does not disambiguate between a specific or a nonspecific interpretation

---

3 Even 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).

of the referents *thaspä ‘apple’* and *šuka ‘dog’*. In these examples, it is the nonspecific article that resolves this ambiguity and marks the two referents as nonspecific. Example (44) illustrates the form *etä* of the nonspecific article in the plural.

(41) [makno’xlova wā] c’el eũ’papi cave *ART:EXSPEC* in laid.s:3pl.o:3sg ‘They laid him in a cave.’ Lakota (*Van Valin 1977: 66*)

(42) [thaspā wâži] wachi apple *ART:NSPEC* want.s:1sg.o:3sg ‘I want an (any) apple.’ Lakota (*Williamson 1984: 48*)

(43) [šuka wâži] ophethu maši dog *ART:NSPEC* buy ask.s:3sg.o:1sg ‘He asked me to buy a (any) dog.’ Lakota (*Williamson 1984: 49*)

(44) [cha etä] aku wa stick *ART:NSPEC.PL* bring IMP ‘Bring back some (any kind of) sticks!’ Lakota (*Williamson 1984: 48*)

In (45), the nonspecific referent of *thaspä ‘apple’* is under the scope of negation. In this case, the nonspecific article bears an additional negation marker and occurs as *wâži-ni* (cf. section 8.2.1).

(45) [thaspâ wâži-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 (45) as a separate negative nonspecific article and list additional exponents for combinations with non-human and inanimate plural nouns. However, examples of nonspecific articles in negation contexts are already rare in the literature; the only marker that is shown in such contexts being *wâžini* (*Ingham 2003, Rood & Taylor 1996, Van Valin 1977, Williamson 1984*). Independently from this issue, I do not treat the nonspecific article exponent *wâžini* in negation contexts as a separate marker because negation itself does not result in the coding of a separate referent type.

*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: 49f.*), which is why I do not treat it as an article here.
5.2.4 The source of nonspecific articles

In section 5.2.1, it was mentioned that the nonspecific article can be traced back to a verbal irrealis marker. This section shows that we find related evidence for the development of an irrealis marker to a nominal nonspecific marker (and eventually, article) in other Siouan and Wakashan languages.

The nonspecific article in Q’anjobal junoq was shown to formally consist of the exclusive-specific article jun and another marker -oq. Even though neither Raymundo Gonzáles et al. (2000) nor Mateo Toledo (2017) mention this nonspecific article, Mateo Toledo (2017) discusses -oq as a verbal irrealis or infinitive marker -oq for intransitive contexts. It is very likely that this irrealis marker -oq fused with jun and developed into the nonspecific article. Examples of the verbal uses of -oq are given in (46) to (48). They show its three uses as irrealis, future, and infinitive marker.

(46) q-q’anjab’ ayach ta q-ach q’anjab’-oq.
   pot-talk to.2 cond pot-2pl talk-IRR
   ‘X will talk to you, if you talk.’ Q’anjobal (Mateo Toledo 2017: 538)

(47) hoq-ach tyaj-l-oq yeikal.
   pot-2 prayer-intr-IRR tomorrow
   ‘You will pray tomorrow.’ Q’anjobal (Mateo Toledo 2017: 538)

(48) x-toj heb’ aw-j-oq b’ay-tu.
   ipfv-3 3pl voice-intr-inf tomorrow
   ‘They went to shout there.’ Q’anjobal (Mateo Toledo 2017: 538)

Although Mateo Toledo (2017) does not present examples in which the article and the irrealis marker co-occur, we find two examples of -oq on nouns, shown in (49) and (50) below. In example (49), the irrealis marker occurs with a noun that functions as the predicate of the sentence. In (50), on the other hand, we see the marker on a noun that in an argument position. It may have a nonspecific referent, as it is in the scope of the question.

(49) man anima-oq hach.
    neg person-IRR 2sg
    ‘You are not a person.’ Q’anjobal (Mateo Toledo 2017: 551)

(50) 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’anjobal (Mateo Toledo 2017: 553)

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 and not with the noun as in the two examples above.
Other unrelated languages show a similar development, the only difference being that the irrealis marker attaches to the noun and not to the article. There is however another example of an irrealis marker that occurs with the indefinite article in a related Mayan language spoken in Mexico, Ch’ol, as is shown in (51). 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.

(51) 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)

A development similar to the one in Q’anjobal is found in Hidatsa (Siouan, USA), only that the development in Hidatsa does not seem to be as advanced as in Q’anjobal. 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 so-called 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 (52), contrasted with the expression of a specific referent in (53):

(52) hiraacá-mià-rúg aru-m-úá-waa-c  
    Hidatsa-woman-COND 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)

(53) hiraacá-mià-wa mi-úá-waa-c  
    Hidatsa-woman-ART:INDEF 1-marry-CAUS.1-DECL  
    ‘I married a Hidatsa woman.’  
    Hidatsa (Park 2012: 368)

Like its Q’anjobal counterpart, the Hidatsa conditional marker *-rúg* usually marks irrealis, conditional, and future on both verbs and nouns. Examples (54) to (55) illustrate its use with a verb and a noun, respectively.

(54) ééhgee-wa-rúg aru-nii-ma-hgi.wéˀ-he  
    know-1.ACT-COND 2.stat-1.ACT-tell-EMPH  
    If I knew it I would tell you.’  
    Hidatsa (Park 2012: 228)

(55) áàda-rúg magi-maa-ihgohbi-wihíˀ-oʾ-’  
    daylight-COND RECIP-1.ACT-meet-FUT.Q.1-PL-Q  
    ‘Shall we meet tomorrow?’  
    Hidatsa (Park 2012: 181)

Also in Crow, another Siouan language, a conditional marker may be related to a marker on nouns that marks them as indefinite (not only nonspecific, in this case). Crow, like Lakota, has an article system with a definite (*-sh*), specific (*-m*) and a nonspecific (*-eem*) article. Examples (88), (89), and (90) illustrate the use of these three article types, respectively.
According to Graczyk (2007: 233), in a certain narrative genre, the two articles -m and -eem from the indefinite domain are replaced by another single marker, -dak/-lak (-dak follows consonants while -lak is used after vowels). The use of -dak is illustrated in examples (59) and (60) below; it no longer distinguishes between specific and nonspecific referents.

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

(60) éehk bal-héelee-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)

Graczyk (2007: 230f) notes that “dak is another indefinite nonspecific determiner; it is homophonous with the conditional and temporal conjunction dak [...] Dak occurs relatively rarely as a determiner. It is best treated as generic irrealis marker that functions both as a determiner and as a complementizer”. While the development of -dak into a broader indefinite marker on nouns is not entirely clear, Crow provides another piece of evidence pointing towards conditional markers from the verbal domain as a source for referential markers in the indefinite domain.

The origin of the nonspecific article wâži in the closely related language Lakota, which formally also corresponds to the specific article wá and another segment -ži, is much less clear. There is not an obvious marker in the verbal domain that would correspond to -ži (/ʒi/). Ingham (2003: 31-32) mentions a negative marker /ʃni/ (spelled -sni) : “The suffix -sni negates predicates and thus only occurs following verbs or nouns when the latter occur as predicates”. However, relating these two markers at this point would only be speculation.

Two more examples of languages that show this “flexibility” of the conditional marker to occur on nouns and result in the nonspecific interpretation are the two Wakashan languages Nuuchahnulth and Makah. Examples (61) and (62) show this for the two languages, respectively.
Thus, there evidence from several languages points towards the development from a conditional or irrealis marker in the verbal domain to a nonspecific marker on the noun, provided that this marker is flexible enough to occur on nouns, or, in the case of Q’anjobal, on the article itself.

5.3 Indefinite articles

Indefinite articles encode specific and nonspecific referents. In this section, I present three examples of this article type: the indefinite article in Tz’utujil (section 5.3.1), Carib (section 5.3.2), and Bonan (section 5.3.3). Although indefinite articles are commonly found in European languages, the choice of examples is intended to show that they are by no means restricted to this area of the world.

5.3.1 Tz’utujil

The marker jun in Tz’utujil (Mayan, Guatemala) is an example of an indefinite article, and was treated as such in Dayley (1985) and Tz’utujil Tinaamitaal (2007). In contrast to the cognate exclusive-specific article jun in Q’anjobal, the article in Tz’utujil is systematically used to mark specific as well as nonspecific referents.

Examples (63), (64), and (65) illustrate the use of jun to mark referents as specific. In (63), the referent of jay ‘house’ is introduced into the discourse; it is a particular referent of its kind set which is not (yet) identifiable by the hearer. In (65), the referent of se’ep ‘gift’ is specific and its identifiability is not relevant in the given discourse situation.

(63) xkatzu’⁰ [jun jay tz’bu’uk’] p’=tinamit
         see.pst:1plo:3sg ART:INDEF house old in=village
     ‘We saw an old house there in the village.’
         Tz’utujil (prim. data)

(64) tetz’ta k’oola [jun ak’aal] chpaam ya’
     look.imp exist ART:INDEF boy inside water
     ‘Look! There is a boy in the water.’
         Tz’utujil (prim. data)
The dance was so beautifully that the grandfather gave a gift to teacher of dancers.

Tz’utujil (prim. data)

The indefinite article jun cannot encode a definite referent. It is the absence of the article, as is shown in example (66b), that leads to a definite interpretation of the referent of k’olbek ’wedding’.

(66) a. ximba p’=[jun k’olbek] go.1sg to=art:indef wedding ‘I went to a wedding.’
    b. ximba p’=k’olbek go.1sg to=wedding ‘I went to the wedding.’ Tz’utujil (prim. 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 (67) and (68), we see that the indefinite article can be used with plural nouns.

(67) 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 (prim. data)

(68) [julee wixb’-ij] xinkeresta’ chaq’a’ iwir art:indef.pl friend.pl.pl came.see night yesterday ‘Some friends came to visit last night.’ Tz’utujil (prim. 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 (69), (70), and (71) illustrate its use with nonspecific referents.

(69) k’ool [jun atz’ib’ab’el]? exist art:indef pen ‘Do you have a (any) pen?’ Tz’utujil (prim. data)

(70) nk’atzin chwa’ [jun atz’aqoneel] need to.1sg art:indef architect ‘I need an (any) architect.’ Tz’utujil (prim. data)

126
Even though examples (69) to (71) show that the indefinite article *jun* is used to mark nonspecific indefinite referents, it is not obligatory and appears to be less natural in at least certain nonspecific contexts involving the existential marker *k’ool*:\(^5\)

Despite the fact that *jun* shows exceptions to its use with referents from the indefinite domain, its use in specific and nonspecific contexts can be assumed to be systematic, which is why I treat it as an indefinite article.

### 5.3.2 Carib

Another example of an indefinite article outside of Eurasia is the indefinite article in Carib (Cariban, Suriname). Examples (74) and (75) show that the marker *amu* is used to introduce new and not (yet) identifiable referents into the discourse. Also in (76) and (78) below, *amu* occurs with referents that are marked as specific because either the unambiguous identifiability of their identity is not relevant in those contexts or neither hearer nor speaker can unambiguously identify the referents.

\(^5\)Neither Dayley (1985) nor Tz’utujil Tinaamitaal (2007) discuss this.
(76) [uwapoto-mpo amu] nirompy-i rapa older.person-dev ART:INDEF die-recpst again
'It happened again that an old person died.' Carib (Courtz 2008: 145)

(77) typo roten [akuri amu] si-ene-ja apart just red.rumped.agouti ART:INDEF 1-see-PRS
'Suddenly, I saw an agouti.' Carib (Courtz 2008: 182)

(78) moro wara ro rapa [woryi mati amu] si-ene-ja y-jena y-woneto that like continue again woman black ART:INDEF 1-see-PRS 1-in.arms 1-dream
'I dreamt and saw a black woman in my arms.' Carib (Courtz 2008: 187)

That *amu* is an indefinite article is shown by its use with referents that can be nonspecific. Examples to illustrate this are given in (79), (80), and (81). Here, neither the article *amu* nor the context disambiguate between a specific and a nonspecific interpretation of the referent.

(79) erome [tonomy amu] si-epory-taka rapa now animal ART:INDEF 1-find-fut again
'Today, I’ll find game again.' Carib (Courtz 2008: 187)

(80) [akuri teràa amu-kon] si-upi-ja reed.rumped.agouti already ART:INDEF-PL 1-seek-PRS
'I am going to try and find some agoutis.' Carib (Courtz 2008: 188)

(81) isème te [amu auto po-no amamin-nano kapy-ry] ‘se wa in.spite.of but ART:INDEF house at-adnom work-nposs make-poss wanting be.1 [amu pyrata-me-mpo] enepy-tòme y-‘wa ART:INDEF money-little-dev bring-instr 1-to
'Still, I want to do some work at home to earn some money.' Carib (Courtz 2008: 198f)

Example (80) shows that the indefinite article is compatible with plural and even bears the only plural marker in the noun phrase.

5.3.3 Bonan

We also find an indefinite article in Bonan (Mongolic, China). Example (82) shows that the indefinite article *-gə* occurs with the noun *ahku* ‘Buddhist monk’ to introduce the nonidentifiable referent into the discourse:
We see the marker -gə with other specific referents in (83) to (86). In (83) and (73), the article occurs in the object position of a transitive clause. Example (85) shows that the indefinite article -gə is also used in the subject position.

Fried (2010: 47) notes that definiteness in object positions is signalled by the absence of the accusative marker -nə, which causes the indefinite article to not be required in this position; a number of specific and/or nonspecific referents in object positions are indeed expressed as bare nouns in Fried (2010). An example is given in (87) below.
Nevertheless, the previous sentences in (83) and (73) showed that the indefinite article does occur in the object position; example (88) shows that -gə can not only be used with objects but that it is compatible with the accusative marker as well.

88  namamada khāl-da=ku  tčačaŋnən jima=ɡə=no  sugə-san=tɕə
long ago speak-IMP=IPFV.NMLZ time goat=ART:INDEF=ACC argue-POS=QUOT.IPFV
khāl-na.
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 for contexts with referents whose interpretation is ambiguous between being specific and nonspecific are given in (74) and (90). We can thus see that a noun marked by -gə can receive a nonspecific interpretation as well.

89  dʒoma [htcəχta ɡə=ɡə]  ap=ku  taraŋ ər-tɕo.
Droma bicycle new=ART:INDEF take=IPFV-NMLZ desire come-IPFV.O
'Droma wants to buy a new bike.'
Bonan (Fried 2010: 82)

90  pə dʒoma=da  samtəχ=ɡə  oχ=ku  taraŋ ər-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 (91a), (91b), and (92) below. Definite referents in Bonan are expressed by bare nouns. In (91a), 'the boxes' are contextually unique in the context of the story told, the same is the case with 'the teachers' in (92) in the context of a story concerning a certain school. The sentence in (91b) features the same referent 'mountain' as (91a), making it an anaphoric referent that is expressed as a bare noun as well.

91  a. səmə=no  ula=da  dabla-tɕə o khar-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: 330)

   b. de  χara-tɕə=ku  tčənəsan sowu  ula=no
   just taunt-IPFV=NMLZ reason important mountain=ACC
   oχ=to=tɕi-san
   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)
Since -gə is systematically used with specific and nonspecific referents and absent in definite contexts, it is an example of an indefinite article.

5.3.4 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 7.2). Nevertheless, this type of article is somewhat different from the other indefinite articles as seen in the previous sections: it mainly occurs with discourse-prominent referents that are introduced into the discourse and correspond to the initial element of a topic chain, or with referents that are human or animate. Since these parameters are highly correlated and are relevant to information structure, I use the cover term “discourse-prominent” here. The most important property of presentational articles is that they seem to be somewhere in the middle of a scale from the numeral ‘one’ with an occasional presentative function on one end and with the indefinite article on the other end. This corresponds to the traditional grammaticalization path for indefinite articles (cf. section 7.2). Therefore, it seems plausible to regard these presentational markers as some sort of emerging specific articles. However, in this section, I propose that a presentational article is not necessarily a marker at an intermediate (and “unstable”) stage between the numeral ‘one’ and the indefinite article.

The first reason for treating these markers as articles is that their use is systematic, even though it is restricted to discourse-prominent referents from the indefinite domain. The second reason is that their use may not be strictly bound to specific referents. Data from a number of languages that I will briefly discuss in this section suggests that presentational articles may also be able to encode nonspecific referents, only that this use can be expected to be highly infrequent given that the article is restricted to occur with discourse-prominent referents. This means that the presentational article is not necessarily in such a middle position between the numeral and the indefinite article, and it means that its restriction to discourse-prominent referents may be, to a certain degree, independent of the restriction to specific referents. The third reason is that regarding presentational articles as emerging specific or indefinite articles implies that they necessarily develop further and extend to the coding of less prominent referents. However, we do not know whether articles that are restricted to discourse-prominent referents are necessarily “unstable” across time; it may also possible that presentational articles never extend to be used with less...
prominent discourse referents. For this reason, I treat them as a separate subtype of indefinite articles here.

One example for what I call presentational articles can be found in Lango. The marker -mɔ́rɔ̂
systematically marks referents as specific, but it only occurs in certain types of contexts. Examples
(93) and (94) below illustrate the use of -mɔ́rɔ̂.

(93) twɔl-ľɔ́  o̭kàò  a̽tín
snake-PRES bite.PERF.3SG child
‘A snake bit the child.’
Lango (Noonan 1992: 162)

(94) pʊ́nɔ́-mɔ́rɔ̂  ti̭̩  í ṃe  ɔ̪́t
pig-PRES be.present.HAB.3SG in back house
‘There’s a pig behind the house.’
Lango (Noonan 1992: 162)

However, -mɔ́rɔ̂ is not generally required to mark referents as specific, as the following two exam-
pies show. Even though the referents of lócə̀ ‘man’ in (95) and mɛ́ ‘girl’ (96) are not identifiable
by the hearer, they do not require a specific marker.

(95) àŋéô  lócə̀  â têk  â ryè
know.HAB.1SG man ATTR strong ATTR wise
‘I know a strong and wise man.’
Lango (Noonan 1992: 164)

(96) nwàŋ  gìn  ɔlärò    pákô  mé  nɔm
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 sen-
tences, are often accompanied by -mɔ́rɔ̂, and describes the distribution of -mɔ́rɔ̂ in indefinite con-
texts as follows:

Not all indefinites are marked with -mɔ́rɔ̂; 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 -mɔ́rɔ̂ (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)

While Noonan (1992) does not further distinguish between specific and nonspecific referents in
the indefinite domain, we do not only find the article -mɔ́rɔ̂ with specific referents as was the case
in the previous examples. Example (97) shows the presentational article in a nonspecific context:

---

7The initial nasal is phonologically assimilated to the last segment of the noun that the article occurs with.
(97)  cèn-nórọ́  pé
ghost-ART:PRES be.absent.HAB.3SG
‘There are no ghosts.’

(Noonan 1992: 147)

This suggests that the presentational article in Lango is used systematically with exactly those referents that are introduced to the discourse and that will stay discourse-prominent. This also correlates with the occurrence of the article in existential constructions. The fact that it is nevertheless used with nonspecific referents shows that it is important to keep its pragmatic extension (from discourse-prominent to less prominent referents) separate from its semantic extension (from specific to nonspecific referent types). Because the article occurs systematically with both types of indefinite referents that are discourse-prominent, I treat -mɔ́rɔ́ in Lango as a presentational article.

Bilua (Bilua, Solomon Islands) is another language with a presentational article. As examples (98) and (99) show, the marker kama/kala (feminine, masculine forms) occurs with specific referents when they are introduced into the discourse and when they are not yet identifiable by the hearer. In addition, both specific referents niabara ‘canoe’ and bazu-bazulao ‘folktale’ stay the topic of the immediately following utterances.

(98)  a.  Se ta ke ere=v=e [kala niabara].
    3PL TOP 3PL make=3SG.M.O=RMP ART:PRES.M.SG war.canoe
    ‘They made a war canoe.’

    b.  Ne=a niabara ta a=daite Bobe=vo.
        PROX.SG.M=LIG war.canoe TOP 1SG=grand.parent Bobe=3SG.M
        ‘This war canoe was my grandfather Bobe’s.’

Bilua (Obata 2003: 79)

(99)  a.  Anga ta a=q=a zari=a v=ai bazu-bazut=o [kala
    1SG TOP 1SG=O:3SG.F=VAL want=PRES O:3SG.M=VAL 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=LIG RED-folktale TOP NEG very true=LIG=3SG.F ...
        ‘…and this folktale, it is not very true, …’

    c.  ...malai silo-silo=a=mu ke=m=ai ibue=k=o
        ...but RED-small=LIG=3PL S:3PL=O:3PL=VAL make.quiet=O:3SG.F=NRFUT
        k=i=ke=ve=ma.
        O:3SG.F=say=3PL=RMP=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 (100) and (101) below.

(100) ...enge ta nge=q=ai zari=a tu k=ov=o [kama 
...[1PL.EXCL TOP 1PL.EXCL=3SG.F.O=VAL want=PRE 3SG.F.O=GET=NOM ART:PRE 3SG.F. 
uri=a=ma saev=o] ... 
good=LIG=3SG.F survive=NOM
‘...we want to get a good life ...’

Bilua (Obata 2003: 93)

(101) Ko=lupao=va, ti ko=noqoe=k=ou ko ore ma tataikili ma 
3SG.F=dislike=PRE and.then S:3SG.F=hold=O:3SG.F=FUT 3SG.F tree or trunk or 
esa [kama pata-pata]. maybe ART:PRE 3SG.F RED-stump
‘If she dislikes it, she will hold a tree, a trunk, or maybe a stump.’

Bilua (Obata 2003: 129)

The two next 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 (102) and (103), respectively, are expressed as a bare noun and do not receive an article.

(102) ...o=marong=a juli=ko 
raki ale. 
...3SG.M=sleep=PRE 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)

(103) reko=a=ma vo=a siele=ko reko ta, tuto 
ale ko 
wife=LIG=3SG.F 3SG.M=LIG dog=3SG.F wife TOP coconut.shell.container in 3SG.F 
i=k=a ... 
put=O:3SG.F=PRE ... 
‘she, the wife, the wife of the dog, she put her food in a coconut shell container ...’

Bilua (Obata 2003: 304)

In Teotepec Chatino (Zapotecan, Mexico), we find another example of a presentational article. Referents that are prominent in the discourse can be marked as nonidentifiable in Chatino by the use of the marker ska’. Examples of its occurrence with specific referents that the hearer cannot identify are provided in (104), (105), and (106) below.8

(104) mn7ą3 me’lo31 [ska’ kto3] ka13 
see.COMPL Carmelo ART:PRE chicken yesterday 
‘Carmelo saw a chicken yesterday.’

Chatino (McIntosh 2011: 106)

---

8The superscribed numbers in the Chatino examples indicate tone, ‘r’ stands for “relaxed tone” (McIntosh 2011: 60).
When she arrived, a woman was there. (McIntosh 2011: 134)

There is a thing on the table. (McIntosh 2011: 102)

The use of the marker $ska'$ is restricted to discourse-prominent referents. With specific referents that are not relevant to the discourse or whose identifiability does not play a role in the discourse, we do not find the marker $ska'$ to mark them as specific indefinite. Examples for this type of referents, expressed by bare nouns, are given in (107) and (108).

Daniel ate an apple. (McIntosh 2011: 94)

'It came out lying on a stone, where it is to this day.' (McIntosh 2011: 117)

Since the use of $ska'$ seems really restricted to referents of typical presentational contexts and other nonidentifiable referents that are prominent within a given discourse sequence, I treat $ska'$ in Chatino as a presentational article. Being a presentational article, $ska'$ is not restricted to specific indefinite referents. The the two examples in (109) and (110) show that $ska'$ also occurs in contexts that seem less typical for presentational articles. In (109), we see that $ska'$ can also be used with a nonspecific referent. Another nonspecific context in (110) on the other hand shows that a nonspecific referent can also be expressed as the bare noun $knya731$.

A person doesn’t live here.' (McIntosh 2011: 104)

Example (111) illustrates another use of $ska'$: it occurs with a nominal predicate, which means that the specific article may also be used in nonreferential contexts. Again, such uses emphasize the importance of separating the two extension processes in the development of articles: the discourse-pragmatic extension to non-topical referents and the semantic extension to nonspecific referents or, in this case, non-referential expressions.
The examples from Lango, Bilua, and Chatino showed that a presentational article as systematic referential marker restricted to discourse-prominent referents does not need to be restricted to specific referents, but can also be used to mark nonspecific referents. This is evidence for the use of presentational articles with discourse-prominent referents being independent from specificity, which is an argument against the seemingly plausible hypothesis of presentational articles being emerging specific articles. It is probably the case that most discourse-prominent referents are specific, which could account for the fact that we mostly find presentational articles with specific referents. The important point is that this is only a correlation without restricting presentational articles to specific indefinite contexts.

Exclusive-specific articles, as was shown in 5.1, are systematically restricted to specific contexts. Although it seems tempting to assume that presentational articles represent an intermediate stage between the numeral ‘one’ and specific articles on the grammaticalization scale (cf. section 7.2), their compatibility with nonspecific referents shows that this is not necessarily the case.

### 5.4 Summary

This chapter discussed the three article types found in the indefinite domain: exclusive-specific, nonspecific, and indefinite articles. I argued that we need to distinguish a subtype of indefinite articles, namely presentational 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 discourse-prominent to less discourse-prominent) development of indefinite articles may correlate but do not necessarily depend on each other. In other words, presentational articles are evidence for the existence of articles which are restricted to discourse-prominent referents but which occur with nonspecific referents. In a way, exclusive-specific articles are the counterpart of presentational articles, since they only occur with specific referents but not with nonspecific ones, without being restricted to discourse-prominent referents only. For nonspecific articles, I showed that they may originate from verbal irrealis markers. The first step in that development is their extension to nominal predicates in irrealis contexts, then to nouns with nonspecific referents in irrealis contexts, and eventually these markers can become systematic nonspecific markers, i.e. nonspecific articles.
Chapter 6

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 (section 6.1) that is used to mark both definite and specific referents. The second article type is the referential article (section 6.2); referential articles mark definite, specific, and nonspecific referents and thus indicate referentiality as such rather than a specific referential function.

6.1 Inclusive-specific articles

In this section, I present three examples of inclusive-specific articles from Bemba, Tongan, and Basque. Inclusive-specific articles co-express definite and specific referents and are thus semantically vague between these two referent types. This article type is not frequently attested across the world’s languages; my sample only contains seven languages with inclusive-specific articles.

6.1.1 Bemba

Nouns in Bemba (Bantu, Democratic Republic of Congo), like nouns in various other Bantu languages\(^1\), 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 (1):

\(^{1}\)E.g. Zulu, Kirundi, Nguni.
\(^{2}\)The 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).
The noun *umuntu* 'man' in (1) 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 an inclusive-specific article in Bemba. Table 6.1 gives an overview of the noun class prefixes and the corresponding article exponents based on Givón (1969: 28f). 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 each singular and plural pairs of what could be treated as 7 different genders. 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 6.1, the augment or what I treat as an inclusive-specific article has different exponents according to the class of the noun.

Table 6.1: Bemba noun class markers

<table>
<thead>
<tr>
<th>Noun class</th>
<th>Article-class</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>u-mu-</td>
<td>mu(u)-</td>
</tr>
<tr>
<td>2</td>
<td>a-ba-</td>
<td>ba(a)-</td>
</tr>
<tr>
<td>3</td>
<td>u-mu-</td>
<td>mu(u)-</td>
</tr>
<tr>
<td>4</td>
<td>i-mi-</td>
<td>mi(i)-</td>
</tr>
<tr>
<td>5</td>
<td>(il)i-</td>
<td>li(i)-</td>
</tr>
<tr>
<td>6</td>
<td>a-ma-</td>
<td>ma(a)-</td>
</tr>
<tr>
<td>7</td>
<td>i-ci-</td>
<td>ci(i)-</td>
</tr>
<tr>
<td>8</td>
<td>i-fi-</td>
<td>fi(i)-</td>
</tr>
<tr>
<td>9</td>
<td>i-N-</td>
<td>ni(N)-</td>
</tr>
<tr>
<td>10</td>
<td>i-N-</td>
<td>ni(N)-</td>
</tr>
<tr>
<td>11</td>
<td>u-lu-</td>
<td>lu(u)-</td>
</tr>
<tr>
<td>12</td>
<td>a-ka-</td>
<td>ka(a)-</td>
</tr>
<tr>
<td>13</td>
<td>u-tu-</td>
<td>tu(u)-</td>
</tr>
<tr>
<td>14</td>
<td>u-bu-</td>
<td>bu(u)-</td>
</tr>
<tr>
<td>15</td>
<td>u-ku-</td>
<td>ku(u)-</td>
</tr>
<tr>
<td>16</td>
<td>/</td>
<td>pa-</td>
</tr>
<tr>
<td>17</td>
<td>/</td>
<td>ku-</td>
</tr>
<tr>
<td>18</td>
<td>/</td>
<td>mu-</td>
</tr>
</tbody>
</table>

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. Therefore, as the
article systematically occurs with nouns in specific contexts as well as definite contexts but is absent with nouns in nonspecific contexts, I treat it as an inclusive-specific article. The contexts themselves are not provided in Givón (1969). I follow his description as well as the translations of the examples. 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 strongly suggests that the article itself does not disambiguate between a definite and an indefinite interpretation and thus co-expresses both referent types.

Two examples to illustrate this are given in (2) and (3). The inclusive-specific article *u*- is obligatory in specific or definite contexts (3a), the lack thereof being ungrammatical (3b):

(2) naalimweene *(a)-ba-ana
   see.pst.1sg ART:INSPEC-cl2-child
   'I saw (some/the) children.'
   Bemba (Givón 1969: 42)

(3) 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
   'The/a man came.'
   Bemba (Givón 1969: 47)

If a nonidentifiable referent occurs within the scope of a negated single past event, it has to be nonspecific. As examples (4) and (5) show, the article cannot occur in such contexts. The referent of *muana* ‘child’ in (4a) contains the article, and is therefore necessarily interpreted as definite or specific. A nonspecific reading is only available in the absence of the augment *u*- as in (4b). Example (5) illustrates this for the referent of *citabo* ‘book’.

(4) 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)

(5) nishiatumine ci-tabo ku-mu-ana
   send.pst.neg.1sg cl7-book cl17-cl1-child
   'I didn’t send any book to the child.'
   Bemba (Givón 1969: 56)

---

3Even though the translation suggests that the referent has to be definite, Givón (1969: 42) marks it as "token" reference. The distinction between "token" and "type" corresponds to the distinction between specificity vs. nonspecificity/genericity (Givón 1969: 41). In this sense, the referent of ‘child’ in (4a) can most probably also receive a specific interpretation.
The article being required in definite as well as specific indefinite contexts and excluded in nonspecific contexts, we nevertheless find systematic gaps, i.e. contexts in which the inclusive-specific article does not occur (Givón 1969). For instance, it does not occur together with a prenominal demonstrative (6), an associative marker (7), or a locative marker consisting of a class prefix (8).

(6) uyu-(*u)-mu-ntu  
    dem-art:inspec-cl1-man  
    ‘this man’  
    Bemba (Givón 1969: 53)

(7) 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)

(8) ali  mu-(*u)-mushi  
    be.3sg cl18-art:inspec-village  
    ‘He’s in the village.’  
    Bemba (Givón 1969: 55)

That the inclusive-specific article is blocked morphosyntactically in such contexts and not for referential reasons can be seen in example (9). In (9a), the inclusive-specific article -u cannot be used because of the presence of the additional class 17 prefix ku-. In (9b), on the other hand, ku is only present on the first of two coordinated nouns, namely on muntu ‘man’. The class 17 prefix ku not being expressed on the noun muana ‘child’, the latter does occur with the inclusive-specific article u-.

(9) 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.’  
    Bemba (Givón 1969: 55f)

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)

6.1.2 Tongan

In Tongan (Oceanic, Tonga), we find another example of an inclusive-specific article. In contrast to Bemba, which only has an inclusive-specific article, Tongan also has a nonspecific article that was presented in section 5.2.2.

Examples (10) to (11) below illustrate the use of the inclusive-specific article (h)e in definite contexts. Example (10) shows the contextually unique referents of fonuá ‘land’ and vakapuna ‘airplane’ that are unambiguously identifiable because there is a single salient plane and land
below the speaker who is flying in the discourse situation. In (11), we see that he is also used with anaphoric referents.

(10) na’a ku sio hifo ki [he fonuá] lolotonga ’eku puna [he
    PST 1sg see down all ART:INSPEC land while POSS:1SG fly ART:INSPEC
    vakapuna]
    airplane
    ‘I looked down to the land while I was flying.’

While contextually unique and anaphoric referents are the two main definite referent types that an inclusive-specific article has to occur with, example (12) below shows that Tongan he can mark spatial deictic referents as well. In this example, the referent of kā ‘car’ is deictic because the context includes a pointing gesture.

(12) ’alu ki mui ’i [he kā]
    go SPAT.P.ALL back LOC ART:INSPEC car
    ‘Go to the back of the car.’

While both Völkel (2010) and Otsuka (2000) mention that the article he can be used with both definite and specific referents, most of the examples these authors provide only show a definite context. However, given the translation in example (13), the article e does not necessarily mark the referent as definite but can also code it as specific. Here, the referent of ’ofa ‘present’ is not relevant or not identifiable by the hearer and is therefore specific.

(13) 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:INDEFSPEC 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.’

---

4Morton (1962) and Churchward (1985) hardly provide any examples of full sentences.
Also Otsuka (2000: 50) mentions that the use of the article \((h)e\) can mark the referent as definite or specific without disambiguating between these two referent types. This is shown in (14). Without further contextual information, the referent of *ta’ahine* ‘girl’, marked by the article \((h)e\) can either be unambiguously identifiable by both the speaker and the hearer and thus definite, or, it can be a single, particular but not unambiguously identifiable referent linked to the expression used, which makes it a specific referent. We see a similar situation in (15); without more contextual information, the referent marked by the inclusive-specific article can be either definite or specific.

(14) na’e sio ’a Sione ki [he ta’ahine]  
\[ \text{pst see abs Sione to ART:INSPEC girl} \]  
‘Sione saw a/the girl.’  
Tongan (Otsuka 2000: 50)

(15) ha’u mo [e afo]  
\[ \text{come with ART:INSPEC fishing line} \]  
‘Bring the/a fishing line.’  
Tongan (Churchward 1985: 25)

In order to be an inclusive-specific article, \((h)e\) must not occur in nonspecific contexts. As was already shown in section 5.2.2, Tongan uses a distinct nonspecific article in such contexts. Example (16) illustrates this again. We see a question with the nonspecific referent of *maa* ‘bread’, which is marked by the nonspecific article *ha*.

(16) ’oku ’i ai [ha maa]?  
\[ \text{prs in there ART:NSPEC bread} \]  
‘Is there some bread?’  
Tongan (Otsuka 2000: 50)

Therefore, we can conclude that the article \((h)e\) is an inclusive-specific article: it systematically occurs with definite and specific referent types without disambiguating between these two, and it is absent in nonspecific contexts.

### 6.1.3 Tepehua

The third example of an inclusive-specific article discussed in this study comes from Tepehua (Totonacan, Mexico). Although it is analysed as a definite article in Kung (2007: 385), I argue in this section that its use in definite as well as specific contexts paired with its absence in nonspecific contexts makes it an inclusive-specific rather than a definite article.

Example (17) shows the use of the inclusive-specific article *juu* with the anaphoric referent of *barda* ‘wall’. In (217), *juu* is used with the contextually unique referent of *laqachaqan* ‘town’, given that there is a single salient town in the discourse situation.
(17) a. 7entons nii paastak-lich juu 7ukxtin nii ka-nawii-ya7 juu
then COMP think-PFV ART:INSPEC boss COMP IRR-make-FUT ART:INSPEC
barda.
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:INSPEC wall
‘Well, they built the wall.’

Tepehua (Kung 2007: 673)

(18) 7an-lich x-st’aa-nta [juu laqachaqan].
go-PFV PST-sell-o:INDEF ART:INSPEC town
‘He went selling in the town.’

Tepehua (Kung 2007: 643)

In addition, the inclusive-specific article can occur together with demonstratives and is thus also compatible with deictic referents, as is shown in (19) below:

(19) [juu 7anu7 x-t’iyun-7an] lapanak waa x-ta-7asaanan.
ART:INSPEC DEM:DIST PST-TWO-POSS:PL people FOC PST-S:3PL-play.instruments
‘Those two people played instruments.’

Tepehua (Kung 2007: 485)

We also find the inclusive-specific 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 (20) below:5

(20) maa xta7amaqpanan [juu papa7-nin juu kaa waa lakak’iwin evid wash.clothes.PST.S:3PL.O:INDEF ART:INSPEC man-PL REL EPIS FOC woods xtat’ajun].
live.PST.S:3PL
‘The men that were living in the woods would wash.’

Tepehua (Kung 2007: 590)

Thus, those previous examples showed that juu in Tepehua is used systematically in various definite contexts. In order to qualify as an inclusive-specific article, juu must also be used systematically with specific referents. Examples (21) to (23) show that this is the case. In (21) and (22), the referents of lhiich’alhkat ‘job’ and serrootii ‘saw’ are particular referents linked to the nominal expressions which are not unambiguously identifiable by the hearer. In these two cases, the identity of the referent is simply not relevant in the given discourse situations. Example (23) on the other hand shows the prototypical specific context in which new participants are introduced.

5Discussing relative structures in Tephua, Kung (2007: 589) mentions that “[t]he relativizer juu is homophonous with the definite [here: inclusive-specific] article juu”. It is plausible to assume that the relativizer juu is not only homophonous but diachronically related to the inclusive-specific article juu.
in the discourse. In all three examples, we see that *juu* is used together with a noun to mark the referents as specific.

(21) t’asa-ni-kan-lich nii ka-xtaq-ni-kan-a7ch [juu lhiich’alhkat].
‘They yelled that they were going to give him a job.’ Tepehua (Kung 2007: 463)

(22) waa ki-jun-ni-li juu liijuuntuu mim-pay nii naa qoxich
FOC O:1-say-DAT-PFV ART:INSPEC deceased POSS:2-father COMP EMPH good
[juu serrootii].
ART:INSPEC saw
‘Your deceased father told me that it was a good saw.’ Tepehua (Kung 2007: 601)

(23) 7alin-li laqa-tam 7a-wilchan maa soq ta-laa-lihitaju
there.is-PFV CL-ART:INDEF CL:ANOTHER-day EVID straight s:3pl-RECIP-find.PFV
[juu 7akumwarii] juu laka-x-chaqa7-7an.
ART:INSPEC friend ART:INSPEC PREP-POS3-HOUSE-POS3.PL
‘One day, two friends met in their houses.’ Tepehua (Kung 2007: 640)

Finally, example (120) below shows that *juu* does not occur with nonspecific referents.

(24) [tam maqaali7] ka-maa-ch’ixtaq-ni-nch juu tuumiin aantu qoxiyaa
ART:PRES rich.person IRR-CAUS-loan-DAT-O:2 ART:INSPEC money NEG good
tuumiin palata.
money better
‘A rich person could loan you money, but it isn’t good money.’ Tepehua (Kung 2007: 616)

6.2 Referential articles

This section discusses referential articles. Referential articles co-express the three major referent types of definite, specific, and nonspecific referents. 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, the referential article in Rapa Nui (section 6.2.1), in Halkomelem (section 6.2.2), and in Baure (section 6.2.3). The article Rapa Nui is an evident example of a referential article; it has a single exponent that is used with all three major referent types. The referential articles in Halkomelem and Baure on the other hand have various exponents that have been analysed as separate articles/determiners in the literature. The markers clearly occur in all three major types of referential contexts, but since their distributions are conditioned by a number of additional criteria that operate along with referentiality, I analyse these markers as different exponents of a single referential article, or as an abstract referential article system that
uses a number of markers to distinguish functions that are related to certain referential functions but cross-cut them.

### 6.2.1 Rapa Nui

Rapa Nui (Oceanic, Chile) has the marker *te* that is distributed across different referential contexts; it is systematically used to mark definite, specific, and nonspecific referents.\(^6\)

Examples (25) to (28) show the article in different types of definite contexts. In (25), the article *te* is shown with an anaphoric referent. Example (26) features a contextually unique referent. The preceding discourse segment introduced a family, setting a scene in which the children go out and bring 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*.

(25) a. He ma’u he oho mai he tu’u, he pu’a he haka kōpiro mo te ntr carry ntr go hither ntr arrive ntr cover ntr caus ferment for art:ref taura mo hiri. rope 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 ntr drag ntr go hither acc art:ref rope ntr arrive hither ntr tie to ruŋa ki te puku ma’ea ena, ’a ka ūŋō rō ka ūŋō rō, ’ā above to art:ref boulder stone med until cntg firm emph cntg firm emph until ka harara rō. cntg 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 2017: 581-582*).

(26) He e’a ia tūŋā pokī era a te vāŋa 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 2017: 576*)

---

\(^6\)In this section, I only discuss the markers *te* and *he* which are used for common nouns. Like in many Oceanic languages, Rapa Nui has a different nominal marker, *a*, that occurs with proper nouns.
In (27), the referent of *nu’u* ‘people’ is deictic since the referent corresponds to the people that are present in the discourse situation; it is thus identifiable by all discourse participants. Finally, (28) shows *te* with an establishing referent.

(27)  
ki a kōrūa [ki te nu’u] hakaroŋo mai ’i a au he ki 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 2017: 46)

(28)  
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  
en a  
‘Come, let’s go to Bethlehem, to see the thing announced (to us).’  
Rapa Nui (Kieviet 2017: 46)

Having shown that the article *te* systematically occurs in various definite contexts, the next contexts in which it is expected as a referential article are specific contexts. This is shown in (29) and (30). In both examples, the nouns are linked to a particular referent from their kind set, but the referents are not identifiable by the hearer. Example (29) shows a typical context in which a new referent is introduced into the discourse, while the specific referent of *henua e hitu* ‘seven islands’ is not unambiguously identifiable by either hearer or speaker. Thus, *te* also occurs with different types of specific referents.

(29)  
i te noho iŋa tuai era ’ā [te taŋata 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 2017: 238)

(30)  
ko tu’u ’ana a au ki ruŋa i [te henua e hitu]  
PERF arrive CONT PROP 1SG above at ART:REF land NUM seven  
‘(In my dream) I arrived on seven islands.’  
Rapa Nui (Kieviet 2017: 238)

In order to be treated as a referential article, *te* is required to mark nonspecific referents as well. Examples (31) to (34) show that this is what we find in Rapa Nui. In these examples, the referents that occur with the article *te* are nonspecific because they do not correspond to any particular referent from their kind that is linked to the noun.

(31)  
e ai rō ’ana hō [te me’e] mo ta’e rova’a e te  
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 2017: 192)
(32) he kiō'oku ki kō'ku nā pokī tainē era mo oho o mātou mo kimī 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 (any kind of) shells.'
Rapa Nui (Kieviet 2017: 239)

(33) ko mate atu 'ana ki [te vai] mo unu
PERF die away Cont to ART:REF water for drink
'I’m dying for (any kind of) water to drink.'
Rapa Nui (Kieviet 2017: 237)

(34) 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 2017: 241)

Given that te occurs with definite, specific, and nonspecific referents, its status as an article and
thus as a primarily referential marker is not uncontroversial. Based on the previous examples
illustrating its use, one may be tempted to argue that it should rather be analysed as a syntactic
marker or as a nominal marker of some sort.

In addition, from a syntactic point of view, te occurs in complementary distribution with the
marker he. The latter is primarily used to mark nouns in predicate position, as is shown in (35).7
Moreover, we see in (36) that he is not felicitous in argument positions in which the article te was
shown to occur in.

(35) [he taŋata] tau manu era
PRED man DEM bird Dist
'That bird was a human being.’
Rapa Nui (Kieviet 2017: 242)

(36) *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 2017: 237)

This makes an 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. In other words, te could be analyzed as
a nominal marker which signals that the word it occurs with as a noun. The marker he, in turn,
would then be a predicate marker, indicating that the word it occurs with is a verb or a predicate.
This is a plausible analysis especially since Rapa Nui has flexible word classes, meaning that many
lexemes can be used as either nouns or verbs. Example (37) illustrates this for the lexeme poki
'(be) child’, which is compatible with both nominal and verbal contexts where it can occur as an

---

7Since a closer look at the nature of the marker he would go beyond the scope of this study, I simply gloss he as
predicate marker following Weber (2003) and Kieviet (2017), although I do not necessarily follow the analysis of he
as a determiner proposed in Kieviet (2017: 244f.).
argument (37a) or as a predicate (37b). We see the same effect for the lexeme *nuinui* ‘big(ness)’ in (38a) and (38b).

(37) a. he pōrekoreko [te ūa pokii] ‘i Tāhāi.
   ntr birth:red art:ref pl child at Tahai
   ‘Children were born in Tahai.’

   Rapa Nui (Kieviet 2017: 76)

   b. mai te hora era ō’oku e pokii 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 2017: 76)

(38) a. [te *nuinui* o Tahiti] e ‘āmui atu ūa e ono *nuinui* nei o Rapa
   art:ref big:red of Tahiti ipfv add away 1du.in num six big:red prox of Rapa
   Nui

   ‘The size of Tahiti altogether is six times the size of Rapa Nui.

   Rapa Nui (Kieviet 2017: 255)

   b. mai ki hāpa‘o nō tātou i a ia ʻātā ka *nuinui* ō
   hither hort care for just 1pl.in acc prop 3sg until cntg big:red emph

   ‘Let us take care of him until he is big.’

   Rapa Nui (Kieviet 2017: 330)

While the article *te* is certainly one of the indicators that *poki* ‘child’ in (37a) corresponds to a noun and not to a verb, also the syntactic position and e.g. number marking indicate the syntactic status of *poki*. Following Kieviet (2017: 238-240) in regarding *te* as an article rather than a syntactic marker, I view the syntactic function of *te* as a secondary one; it is rather a consequence of its distribution as a referential marker, since referential expressions occur in argument positions while predicates express events and not discourse referents.

Furthermore, an analysis as a referential article accounts for the contexts *te* occurs in, and also for those it does not (and in which the marker *he* is used instead), while a syntactic only accounts for a subset of contexts in which *te* does not appear. In addition to predications, the article *te* cannot be used in the following contexts: with the instrumental preposition *hai* ‘with’ (39), with the comparative preposition *pē* ‘like’ (40), and in appositions (41). While no other prenominal marker occurs after the preposition *hai* in (39), examples (40) and (41) show that *he* is (can be) used with the noun instead.

(39) he tunu mā‘ea vera haka hopu i te pokii [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 2017: 264)

---

8For other minor constructions that condition the distribution of *te* and *he*, cf. Kieviet (2017: 235f.).
‘Small like a mouse, guarding the house like an insolent dog.’

Rapa Nui (Kieviet 2017: 244)

‘The father of Te Rau, the leader of the people of Kapiti, said …’

Rapa Nui (Kieviet 2017: 243)

The absence of te in these contexts can be accounted for by the conventionalization of te marking referential nominal elements. Especially for nouns that express instruments or that occur in simulative constructions, we can assume that they are typically not used to refer but rather to evoke a particular concept. In appositions, the referent is usually identical to the referent in 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 blocking effects for the article. This can account for the incompatibility of the referential article te in Rapa Nui with the constructions shown in (39) to (41). While te cannot occur with the prepositions shown in (39) and (40), the use of te is by no means generally blocked in the presence of prepositions. As we can see in (42), te occurs with the spatial preposition i ‘at’. Expressions with spatial referents are not typically non-referential, which means that their use together with a referential article would be expected.

A purely syntactic account captures parts of the distribution of te, but it would seem less evident why te cannot occur with certain prepositions, while it is required with others if it was a nominal or argument marker. Hence, I treat te as a referential article whose referential properties can also account for its syntactic behaviour.
6.2.2 Halkomelem

Musqueam\(^9\) Halkomelem (Salishan, Canada) has a number of markers 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 are better regarded as exponents of a single referential article.

An overview of the article exponents in Halkomelem based on Suttles (2004: 339) is given in Table 6.2.\(^{10}\) In addition, the article has different exponents with regard to a category of deixis and of visibility. I argue in this section that the marking of deixis is only an additional function of the referential article. Table 6.2 shows the exponents of the article according to its deictic functions. Suttles (2004: 340) distinguishes between proximal (he calls it “present”) and visible, proximal (“nearby”) and nonvisible, and remote.\(^{11}\)

<table>
<thead>
<tr>
<th>Article Exponents</th>
<th>Proximal &amp; Visible</th>
<th>Proximal &amp; Invisible</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tə</td>
<td>kʷθə</td>
<td>ḱʷə</td>
</tr>
</tbody>
</table>

The following examples illustrate the deictic functions of the article. In (43) the referent of sqʷəméý ‘dog’ is marked as proximal and invisible by using kʷθə; 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 (44) with the article tə marking visibility corresponds to a scenario in which the speaker again is outside the house, and in which the dog is coming outside towards the speaker.

(43) niˀ skʷtéxʷ tə lëləḿ [kʷθə sqʷəméý]
be.there inside OBL ART:REF.NVIS house ART:REF.NVIS dog
   ‘The dog is in the house.’ Halkomelem (Suttles 2004: 342)

\(^9\)Halkomelem can be split into three major groups of dialects: Island varieties, Downriver varieties, and Upriver varieties. The Musqueam variety of Halkomelem belongs to Downriver varieties of the language. Although I only refer to the Musqueam variety in this section, the determiner system seems to be similar in Island Halkomelem (e.g. Gerdts & Hukari 2004). The “determiner system” in Upriver Halkomelem, on the other hand, shows greater differences, for details see Brown & Thompson (2013), Galloway (1993), Wiltschko (2002).

\(^{10}\)Suttles (2004: 339-340) further distinguishes “feminine” forms of the articles; however, he also notes that these are used only with human and female referents. For this reason, I do not discuss these forms here.

\(^{11}\)Suttles (2004: 340) lists additional shortened variants of the article exponents. Since these are rare in the examples that he provides, I do not list them here. In addition, (Suttles 2004: 348) treats another marker as an 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.
(44) ʼi ʼəmí ᵁXqəl [tə sqʷəméy]  
be.here come exit ART:REF.VIS dog  
‘The dog came out.’  
Halkomelem (Suttles 2004: 342)

Example (45) shows that the marking of deixis and visibility is not always straightforward: the article *tə* for visible proximal referents is rather contrasted with the “remote” form *kʷθə* instead of being contrasted with the proximal invisible form *kʷθə*. Another example for the deictic use of *kʷθə* is given in (46), although Suttles (2004: 344) notes that it additionally implies that the hearer cannot identify the pasture.

(45) a. niʼ ʼə [tə nə-ʼéθəqən]  
be.there OBL ART:REF.VIS my-front  
‘in front of me (in a canoe)’ literally: ‘there at my front’

b. niʼ ʼə [kʷə nə-sliʼáaqʷı́]  
be.there OBL ART:REF.REM my-rear  
‘behind me (in a canoe)’ literally: ‘there at my rear’  
Halkomelem (Suttles 2004: 344)

(46) néḿ cən ʼə [kʷə spélxən]  
go I OBL ART:REF.REM pasture  
‘I’m going to the pastures (way off, out of sight).’  
Halkomelem (Suttles 2004: 344)

Although we saw in the preceding examples that the article in Halkomelem is involved in marking deixis and visibility, they are better accounted for as different exponents of a single referential article that systematically occurs with definite, specific, and nonspecific referents and is absent in nonreferential contexts.¹²

Definite referents that are marked by the article are presented in examples (47) to (49). Example (47) shows the first mention of *smə́yəθ* ‘deer’ in a story about deer hunting. In the context of the story, the referent is contextually unique and thus definite. An anaphoric referent is shown in (48). The sentence that it occurs in follows the sentence in (47). Example (49) features a sentence that is uttered with the dog being present in the discourse situation; therefore, the referent of *ʼənsqʷəméy* ‘your dog’ is deictic. These examples also show that visibility is coded independently from definiteness: in (47) and (48), the form *kʷθə* is used to mark nonvisibility, while example (49) shows the form *tə*, encoding visibility.

---

¹²Gerdts & Hukari (2004: 159) make a similar remark, noting: “Semantically, articles refer to definite and indefinite NPs indiscriminately, as is general in Salish languages.”
‘They will be the ones who will shoot the deer when they start coming through.’

Halkomelem (Suttles 2004: 525)

‘...when they come chasing the deer.’

Halkomelem (Suttles 2004: 525)

‘Does your dog bite?’

Halkomelem (Suttles 2004: 471)

The following examples illustrate the use of the article to mark specific referents. Again, we see that the deixis parameter of the article exponents is independent of the referential value. Thus, the form \( t\a \), indicating visibility (perhaps due to the reported visibility), is used to express the specific referent of \( čičí’qən ‘mink’ \) in (50). Another example of a specific referent is provided in (52); the specific referent is marked by the form \( k^\nuθə \) of the article.

‘Then we saw a mink.’

Halkomelem (Suttles 2004: 347)

‘She took (some) pitch.’

Halkomelem (Suttles 2004: 348)

‘...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 (\( k^\nuθə \)). Examples of nonspecific referents are presented in (53) to (55). Note that the article surfaces as shortened \( k^\nu \) in (54). Example (55) shows that also nominalized predicates, here \( nə-s-nəm ‘my going’ \) can be referential and are marked by the article accordingly.

‘I want some (any kind of) pie.’

Halkomelem (Suttles 2004: 345)
DRAFT

We’d better go look for (any kind of) flounders offshore.’ Halkomelem (Suttles 2004: 345)

I want to go.’ literally: ‘What I want is my hypothetical going.’ Halkomelem (Suttles 2004: 345)

Examples (56) to (59) illustrate the lack of the article with non-referring nouns. In (56) to (57), the nouns are used as predicates and therefore have no referent. In example (59), the expression *təw-cɪtməxʷ* ‘owl’ is used in a simulative construction and is non-referring as well.

In contrast to nominal predicates, the article occurs with nouns in equational constructions. This is shown in (60).

Halkomelem (as is commonly attested in Salishan languages (cf. Beck 2013)) has flexible word classes, which means that many words can occur as nouns or as predicates. I do not discuss the status of nouns and verbs in the grammar of Halkomelem here; for the purposes of this section, suffice to say that words which are typically referring can equally be used as predication and vice versa. Given that there is no strict distinction between words used for reference or predication in Halkomelem, we would also expect words that are typically predicates to occur as a referring
expression. Examples (61) to (63) show that this is what we find, and that the article occurs with such expressions to mark them as being referential.

(61) nə́wə yaxʷ ce’ xá’łəm-ət [tə háýqʷ] 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)

(62) stém kʷə [tə kʷəné-t-əxʷ]
what then ART:REF.VIS hold-TR-you
‘What is it you are holding?’ literally: ‘What is your holding?’

Halkomelem (Suttles 2004: 385)

(63) stém [kʷə ’ə-s-cƛ́í’]
what ART:REF.REM you-nmlz-want
‘What do you want (e.g. of food laid out)?’

Halkomelem (Suttles 2004: 400)

To conclude, this section showed that the three markers tə, kʷθə, and kʷə can mark distinctions of deixis and visibility, but it also showed that they systematically occur with different types of referents, and that they are systematically absent in nonreferring contexts. Therefore, I treat them as exponents of a single referential article.

6.2.3 Baure

Baure (Arawakan, Bolivia) features what is described as a system of determiners in Danielsen (2007: 310-316). In this section, I argue that the different markers should rather be analysed as a referential article, similarly to the referential article in Halkomelem.

The system of determiners, following the labels of Danielsen (2007: 311), consists of the following markers:13

| Table 6.3: The system of determiners in Baure as presented in Danielsen (2007: 311) |
|-------------------------------------|------------------|------------------|
| singular                           | masculine        | feminine         | plural |
| article                            | to               | to/ti            | to     |
| present demonstrative              | te               | ti               |        |
| proximate demonstrative            | teč              | tič              | to neč |
| distal                             | ten              | tin              | to nen |

13 Only the article with the exponent tə is glossed as ART in Danielsen (2007). The three demonstrative series are glossed as DEM1, DEM2, DEM3, respectively. Arguing that the markers in Table 6.3 are all different exponents of a single referential article, I gloss them as ART:REF here.
Danielsen (2007: 311) distinguishes four different determiners that have different exponents in the singular and the plural, and for masculine and feminine nouns in the singular. The four determiners are what she calls the article and three different series of demonstrative markers: present, proximate, and distal. Danielsen (2007: 311) notes that “phonologically all determiners seem to be related” and that “[the] exact meaning [of the determiners] is often unclear”. While the functions of the determiners are slightly different but show a great deal of overlap, especially with regard to their referential functions, I argue that they should rather be treated as different markers that belong to the same, abstract, referential article. In addition, the paradigm shows syncretisms between the so-called article and the present demonstrative for feminine and plural forms, which makes the so-called article even less distinguishable from other other demonstratives, given that in a number of contexts, their form and function overlap.

According to Danielsen (2007: 312), the functions of to (or what she calls article) are the following:

“The article does not mark definiteness. It is generally used with proper names and heavenly bodies, such as to ses ‘the sun’ or to kiher ‘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.”

While it may not be the case that to introduces new salient referents into the discourse, it occurs with definite, specific, and nonspecific referent types. Examples of its use in definite contexts are given in (64) to (77), with an anaphoric and a contextually unique referent, respectively.

(64)  teč  worapik teč  ses  ro=aceroko-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)

(65)  ro=kic0-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)

Examples (66) to (68) illustrate that to does in fact occur with referents that are specific, because their link to the expression is not unambiguously identifiable by the hearer.
(66) 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)

(67) 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)

(68) 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 i.e. there is no single referent that would be linked to the expression. The use of to in such contexts is illustrated in examples (69) and (70) below:

(69) 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)

(70) no=ačo-no-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, amongst others nominalizations. In (71) and (72), we see two examples that illustrate the use of to marking a verbal expression as a referential expression.

(71) 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)

(72) 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 (73) and (74).
(73) a te kač moro’in=no 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)

(74) 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 show that we have a functionally versatile marker that, on the one hand, marks a nominal expression as a referential one, and occurs in different fixed constructions on the other hand. The latter may suggest that to is not necessarily a referential marker but rather a nominal marker of some sort. However, it holds for to in a similar way to the referential article te in Rapa Nui, that its property to mark an expression as a nominal one rather follows from its primary function of marking an expression as a referential one.

That what is treated as an article in Danielsen (2007) can formally overlap with what she treats as demonstratives was shown in Table 6.3 in the beginning of this section. In order to show that all these markers should indeed be treated as exponents of a single referential article, I show in the following parts of this section that the so-called demonstratives cover similar functions as to. Even though there are different preferences and minor differences for their uses, there is no clear-cut difference in their functions, especially regarding referentiality. Nominal expressions that are referring require one of these determiners: “When an NP is used referentially, a determiner is generally obligatory” (Danielsen 2007: 310). Nominal expressions that function as predicates do not require the presence of a determiner. Because of that, the main common function of all the determiners is arguably the indication of referentiality. Each determiner may add a certain (discourse) pragmatic effect in certain context which might account for the variation that we find.

The so-called present demonstrative, te/ti (in the singular), is described as follows in Danielsen (2007):

“The present demonstratives are used like the article for topic NPs, but they are differentiated for gender (which has partly been merged with the article as well). As demonstratives they have a less pointing function than teč/tič ‘DEM2M/F’, but they are generally used in contrasts […] The demonstratives te/ti ‘DEM1M/F’ are called “present” because they are generally used with something present, maybe visible, already mentioned or topic.” (Danielsen 2007: 312)

Example (75) shows te in an anaphoric context, which is expected given that they are analysed as demonstratives and characterized to occur with anaphoric and discourse prominent, topical referents. Based on the text collection provided in Danielsen (2007), the form te/ti however does seem less frequent than teč/tič with anaphoric and topical referents.
(75) a. heni teč sipori ro=kič-wo teč kotis. yes ART:REF frog 3SG.M=say-COP ART:REF lizard
    ‘“Yes”, said the frog to the lizard.’

b. koeh(koe’) [te kotis] ver ro=eheipoeko-wo koeč ver so.that ART:REF.M lizard PERF 3SG.M=fall.down-COP because PERF
    ro=mani-wapa ač ro=siapo noiy ro=peni-ye. 3SG.M=be.cold-cos and 3SG.M=enter there 3SG.M=cave-LOC
    ‘Because the lizard fell down because of being cold and he entered his cave there.’

Baure (Danielsen 2007: 452-453)

Even though te/ti mostly occurs with definite referents, examples (76), (77), and (78) show that it can also occur with referents that are specific and thus not (yet) identifiable by all discourse participants:

(76) ikiy-i-ye te rasroe’ kwe’ [te kokon].
    middle-CLF:FRUIT.BIRD-LOC ART:REF.M orange exist ART:REF.M worm
    ‘There is a worm in the middle of the orange fruit.’ Baure (Danielsen 2007: 199)

(77) kwe’ [ti eton] pari-ye.
    exist ART:REF.F woman house-LOC
    ‘There is a woman in the house.’ Baure (Danielsen 2007: 206)

(78) ro=ino-wo [te kotowor].
    3SG.M=resemble-COP ART:REF.M buzzard
    ‘It (the bird) looks just like a buzzard.’ Baure (Danielsen 2007: 215)

The next marker, the so-called proximate demonstrative teč/tič/to can be used for spatial deixis. Example (79) shows “[…] this demonstrative used emphatically with a pointing gesture […]” (Danielsen 2007: 314):

(79) nti’ ni=yok [te ka’an].
    1SG 1SG=sting ART:REF.M animal
    ‘I will kill this animal with an arrow’ (Danielsen 2007: 314)

Besides its deictic function, Danielsen (2007: 313-314) characterizes the functions of teč/tič/to as follows:

“Proximate demonstratives are used for introducing a character, which may then also be referred to by the same throughout the whole narration. […] the demonstrative is widely used like an article and not necessarily deictically. It may also be related to foregrounding, as it is generally attached to the main characters throughout in a narration. […] The demonstrative
pronoun can also be used anaphorically, referring to an afore mentioned item." (Danielsen 2007: 313-314)

In the texts provided in Danielsen (2007), the marker teč/tič/to occurs relatively often with anaphoric referents; example (80) illustrates this function:

(80) ač hepčin  ri=ko-viano-wo  to  it tiow teč
and it.seemed  3sg.f=attr-companion-cop art:ref eel  cleft art:ref.m
ri=veyono-wo  [teč it].
3sg.f=make.love-cop art:ref.m eel
‘And it was the case that she was companion to the eel; she was the one who made love with the eel.’

Baure (Danielsen 2007: 454)

The function of teč/tič/to to introduce new (specific) referents into the discourse is similar to the use of the English demonstrative this in such contexts. Examples (81), (82), and (83) illustrate its use in establishing contexts. The two referents of eton ‘woman’ and mos ‘mother-in-law’ in (81) are established as identifiable discourse referents in (81a) and continue to be discourse topics in (81b). In (82), the identity of the referent erosorekočonev ‘lassoes’ is established in the following relative clause; the same holds for the referent marked by kori ‘arrow’ in (83).

(81) a. nakiroko-ye kwe’ [tič eton] ač kwe’ [tič ri=mos]
long.ago-loc exist art:ref.f woman and exist art:ref.f 3sg.f=mother.in.law
napiri’ noiy.
also there
‘Once upon a time there was a woman and her mother-in-law.’
b. ač neriki [tič ri=mos] kač ri=epha=ri.
and now art:ref.f 3sg.f=mother.in.law dir 3sg.f=spy=3sg.f
‘And now the mother-in-law went to spy upon her.’

Baure (Danielsen 2007: 313)

(82) nti’ ni=hinoeko-wo [teč ni=erosorekočo-nev] no=wohik=ro.
1sg 1sg=look.for-cop art:ref.m 1sg=lasso-pl 3pl=steal=3sg.m
‘I am looking for my lassoes that they have stolen.’
Baure (Danielsen 2007: 404)

(83) ro=kotiro-wo [teč ro=kori] ro=yoko-wo to areno-nev-či.
3sg.m=have-cop art:ref.m 3sg.m=arrow 3sg.m=pierce-cop art:ref bird-pl-dim
‘He had an arrow that he was shooting little birds with.’
Baure (Danielsen 2007: 204)

Danielsen (2007: 313) relates the foregrounding effect of teč/tič/to neč to an interesting interaction with the forms ti/te: in structurally ambiguous clauses, the argument marked by teč/tič/to neč tends to be interpreted as the agent, while the one marked by ti/te is more frequently associated with the patient.
The final series of demonstratives mentioned in Danielsen (2007) are the so-called distal ones. She describes their use as follows:

“The demonstratives ten/tin/to nen have been subsumed under the term distal, not necessarily interpreted in local terms. It can also mean disapproval and absence. When the distal demonstrative is used with an NP, it generally moves the argument further away from the speaker and hearer […] The distal demonstrative is often used with the negative connotation of disapproval […] and demonstrative is used frequently with arguments absent at the moment of speaking.” (Danielsen 2007: 315)

An example of ten with a definite referent that is absent from the discourse situation is given in (84) below. Danielsen (2007: 316) describes the context of the utterance as follows: “Example [(84)] was uttered when I was looking at unripe plantains, wanting to eat one that day. The speaker went to fetch a ripe one, saying this before she left.”

(84) ač kwe’ [ten pon yi-yakon].
    and exist ART:REF.M other INT-ripe
    ‘And there is another really ripe one.’ Baure (Danielsen 2007: 316)

Its other function of moving the argument further away from the speaker seems to correspond to the conventionalized construction of to nen ‘those who’ in establishing contexts. In that, the Baure construction with to nen … is a relatively close mirror image of the German or English establishing constructions diejenigen, die … and those who … that do typically not alternative with other demonstratives (e.g. ?diese, die … or ?these who …). Examples of to nen in establishing contexts are given in (85) and (86). Another example with tin establishing a singular referent in the discourse is given in (87) below.

(85) koeč to nen ikomoriko-no-wo te howe’ moeh no=hinok=ro.
    because ART:REF.PL kill-NMLZ-COP ART:REF dolphin CERT 3PL=see=3SG.M
    ‘Because those who kill a dolphin can see it.’ Baure (Danielsen 2007: 316)

(86) ro=hinoko-po-wo=hi to nen worapik arenono-a-pik.
    3SG.M=see-PFV.REFL-COP=QUOT ART:REF.PL already.come bird-NMLZ-LNK-COME
    ‘He (only) looked that those who came flying like a bird.’ Baure (Danielsen 2007: 315)

(87) neriki ne’ te yi=kepi-yo-wo kwe’ ha kwe’ [tin marip] niko-no
    now here ART:REF 2PL=Speak-LOC-COP exist HESIT exist ART:REF.F witch eat-NMLZ
    činti-nev.
    person-PL
    ‘Now in that place that you are talking about, there is a witch that eats people.’ Baure (Danielsen 2007: 316)
Summing up the functions of the different Baure markers addressed in this section, two important things have to be noted. The marker *to* occurs with definite, specific, as well as nonspecific referents and could thus be classified as a referential article. On the other hand, the text collections in *Danielsen (2007)* show that while referential nominal expressions systematically occur with a determiner, this is by no means always *to* but often one of the so-called demonstratives are used instead. This makes the status of *to* itself as referential article somewhat questionable, given that other markers are often used at the expense of *to*. What is labelled as “present demonstratives” in *Danielsen (2007)* can mark definite (anaphoric) and specific referents as well and appears to be mostly used to mark anaphoric referents that were not mentioned during a longer discourse segments and that are re-introduced. The so-called “proximate demonstratives” correspond to the determiner that is most often used to mark anaphoric referents as definite in the text collection in *Danielsen (2007)*. In addition, also this series of markers was shown to introduce new, specific referents into the discourse, mostly those that become the main discourse topic in the immediately following discourse segment. The last demonstrative series distinguished in *Danielsen (2007)*, the so-called *distal* ones, occur with definite referents and are often used to establish discourse referents; especially the plural form *to nen* appears to correspond as a fixed construction for establishing contexts.

Hence, *to* and the other markers from the three demonstrative series may have specialized functional domains to a certain extent, but we saw that they share most of their functions in terms of referential domains in which they systematically occur. In addition, it is not only that their referential functions overlap; as was shown in the beginning of this section in Table 6.3, their exponents overlap formally. A clear formal distinction of the four different series of markers is only available for singular masculine referents; the feminine and plural forms do not make a strict distinction between the so-called article and the so-called present demonstrative. On the one hand, one can argue that this fact makes it difficult for the linguist to properly distinguish between those two types of markers. More importantly, we can assume that neither do the speakers necessarily conceptualize them as exponents of two separate markers. Because of this functional overlap regarding referential uses, the high degree of variation that could be idiosyncratic or due to additional pragmatic factors, and because of the formal overlap that makes it difficult to speak of separate markers to begin with, I treat all the markers discussed in this section as exponents of a single, abstract referential article.

In order to argue that we deal with a referential system in Baure, examples (71) and (72) showed that the marker *to* could occur with verbal expressions and mark them as referring expressions rather than as events. The same can be said about the other determiners (*Danielsen 2007: 310*).

---

15For instance, *Danielsen (2007)* glosses all occurrences of *ti* as a demonstrative and never as an article.
Examples (88), (89) and (90) show the use of *te, ten* and *teč* with nominalized and thus referential verb forms:

(88) po-morkoe-š-wapa [te ni=kotive-wo].
    one-year-one-cos ART:REF.M 1SG=be.ill-cop
    'It is already one year now that I have been ill.' Baure (Danielsen 2007: 185)

(89) ač kwe’ [ten pon yi-yako-no].
    and exist ART:REF other INT-ripen-NMLZ
    'And there is that other ripe one.' Baure (Danielsen 2007: 189)

(90) hepčin tiow tič eton maspoen ač ri=hewe-sa-wo [teč it.seemed CLEFT ART:REF woman crazy]
    and 3SG.F=throw-CLF:water-cop ART:REF
    ri=maspoe-pi].
    3SG.F=be.crazy-NMLZ
    'It was the case that the woman was crazy and jumped into the water because of her craziness.'
    Baure (Danielsen 2007: 458)

As a referential article, nouns that are not referring should systematically occur without it. This is what we find in Baure: examples (91), (92), and (93) show three such contexts.

(91) 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)

(92) 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)

(93) puhhh ver howe-wapa=ri.
    puhhh PERF dolphin-cos=3SG.F
    'Puhhh, and she changed into a dolphin.'
    Baure (Danielsen 2007: 195)

### 6.3 Summary

This chapter presented of two article types that crosscut the definite and the indefinite domain: inclusive-specific and referential articles. The former are used to mark definite and specific referents, while referential articles do not mark certain referent types as such but rather indicate that the nominal expression is a referential one; they occur with definite, specific and nonspecific referents. I showed a more evident example of a referential article in Rapa Nui, consisting of a single
exponent that is used in all relevant contexts. The two other examples, the referential articles in Halkomelem and Baure, illustrated a reanalysis of a complex determiner system consisting of a number of different exponents as systems of a single referential article. I argued that we should treat those different determiners as exponents of a single referential article, since referential contexts systematically require a nominal expression to be accompanied by one of the determiner exponents. Moreover, while each of the exponents could be shown to have certain additional properties that the other ones in the system did not have, all exponents showed a great deal of functional overlap, especially with regard to their referential function. Thus, the exponents could not be distinguished on the basis of different referential functions and they only showed different additional pragmatic or deictic functions in certain cases, which is why I presented them as different exponents belonging to a referential article.
Chapter 7

The distinction between articles and similar categories

It is difficult to delimit certain types of articles from other similar elements, especially if the articles develop from these. The three major cases that I will discuss in this chapter concern definite articles and demonstratives, indefinite and exclusive-specific articles and the numeral ‘one’, as well as nonspecific articles and negative polarity items (NPIs). The former two cases are due to the grammaticalization of demonstratives and the numeral ‘one’ into definite and exclusive-specific/indefinite articles, respectively. Their development being a gradual process, it is difficult to motivate a clear cut-off point between the source elements and articles. Therefore, I propose criteria for a distinction between definite articles and demonstratives in section 7.1, and for indefinite/specific articles and the numeral ‘one’ in 7.2. With respect to nonspecific articles and NPIs, the similarity is not due to a diachronic relation. However, since the nonspecific article is not a well-known type of articles, I will point out the differences between this article type and NPIs in section 7.3.

7.1 Demonstratives

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 (1978b: 246), and was revisited in various later studies (De Mulder & Carlier 2011, Diessel 1999, Leiss 2000, Lyons 1999). 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 (1978b: 246), as is shown in Figure 7.1:
Relevant here is the development between stage 0 demonstratives and stage I definite articles. As Schroeder (2006: 554) notes, the contexts of (spatial) deixis, anaphora, and uniqueness of the referent can be viewed as a continuum on a scale from demonstrative to article. 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), i.e. the encoding of anaphoric referents. 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 (or text deixis) as “anaphora”. To be correct, this use of deixis is not completely restricted to spatial deixis but refers to physical perception and attention drawing in a broader sense.

Through this extension of the deictic demonstrative to anaphoric, and at a later stage also other, uses, the marker may become more frequent, which in turn leads to a weaker association of the marker with deictic referents, until its function includes the systematic encoding of situationally unique referents, making it a definite article. Due to this strong connection between demonstratives and definite articles, it is not always clear how to classify markers that are somewhere in between a “typical” demonstrative and a “typical” article situated on the two ends of a scale. The traditional cut-off point between demonstratives and articles seems to be between the marking of anaphoric and situationally unique referents. Lyons (1999: 54) notes: “[…] it may be unclear whether a determiner specialized in anaphoric use is an article or a demonstrative”. Also in other studies concerned with demonstratives, definite expressions, or articles, anaphoric markers are usually treated as demonstratives rather than as articles (e.g. Diessel 1999, Himmelmann 1997, Lyons 1999).

Assuming that the development of definite articles from demonstratives involves their functional extension from markers of deictic referents to anaphoric, and finally situationally unique referents, we can expect to find markers that encode different combinations of functions along this scale. Figure 7.2 shows the functional domains of such markers.1

---

1(Himmelmann 1997: 93-97) proposes a different path for the development of definite articles from demonstratives through establishing uses. Since I only make use of these main functions and their scalar arrangement that is synchronically reflected in the functional domains of different types of demonstratives and articles in the world’s
Breaking down the relevant functions to deixis, anaphora, and uniqueness, we usually find these five combinations of their expression. As is shown in Figure 7.2, I consider markers as demonstratives if they are restricted to expressing deixis, and if they encode both deixis and anaphoric referents (type I and II). As was mentioned above, anaphoric markers (type III) are often treated as demonstratives rather than as articles (e.g. De Mulder & Carlier 2011, Diessel 1999, Dryer 2013b, 2014, Himmelmann 1997). Although Himmelmann (1997: 66) also discusses anaphoric markers on a par with other demonstratives, he notes that this categorization of an anaphoric marker in Wubuy (Nunggubuyu, Australia) might be problematic, as it cannot be used to mark spatial deixis and thus differs considerably from the functions usually ascribed to demonstratives.

Primarily concerned with the definite article in English, Hawkins (1978) defines demonstratives as markers of spatial deixis (“immediate situation use”) and anaphoric uses (“direct anaphoric use”). The demonstrative contrasts in this way with the definite article in English, which is used in different definite contexts based on the uniqueness of the referent. He distinguishes between two types: bridging (“associated anaphora”) and (situational) uniqueness (“larger situation use”). While this distinction may be sufficient for English, it cannot be applied to languages with markers that encode different combinations of functions.

Thus, in order to compare articles across the world’s languages, I treat markers that only encode anaphoric referents (type III) as well as markers that encode both anaphoric and situationally unique referents (type IV) as articles, as long as they comply with the other criteria discussed in Chapter 2. Markers of type V probably correspond to the most common type of definite articles and their status as definite articles is unproblematic. This is not the case with anaphoric markers of type III. Therefore, the following paragraphs address the question concerning anaphoric mark-
ers being treated as articles, which requires them to be sufficiently distinct from demonstratives. The prototypical deictic use of demonstratives can usually be accompanied by a pointing gesture (Diessel 2013b: 243). This link between pointing gestures and demonstratives is well known in the linguistic literature (e.g. Brugmann 1904, Bühler 1934, Clark 1996, Eriksson 2009, Levinson 2004) and was often tied to the spatial deictic function of demonstratives. Diessel (2013b: 243) points out that the function of such demonstratives is not only spatial location, but more importantly, drawing the hearer’s attention towards another 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, 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 2013b: 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 2013b: 245). A purely anaphoric marker is different. Whether or not one wants to argue that it represents a more abstract, extended, or grammaticalized function, an adnominal anaphoric marker is clearly less basic in the sense that it is not present in all languages. Most often, it is the deictic demonstrative that can be functionally extended to mark anaphoric referents. For this reason, I treat markers that only encode anaphoric referents as articles here.

Entirely deictic demonstratives (type I) have to be infelicitous in anaphoric contexts based on their definition. It is not clear whether such demonstratives are attested at all in the world’s languages: even with a distinct anaphoric marker existing in the language, they often seem to be applicable in certain anaphoric contexts. Nevertheless, we find contexts in which the deictic demonstrative cannot be used to mark an anaphoric referent. This is illustrated with examples from Tikuna (Tikuna-Yuri, Colombia) and Limbum (Bantu, Cameroon) below. Examples (1) and (2) below show the Tikuna demonstrative in deictic contexts. The same demonstrative cannot be used to mark (at least certain types of) anaphoric referents, as we can see in (3):

(1) nükü dau [jema putüra] na mëeche
s:2sg.o:3sg see DEM flower cop beautiful
‘Look, this flower is beautiful.’ Tikuna (prim. data)

(2) [jema buun] na yaichi inekúwa nucha dau iranichi
dEM child 3sg.m grow.up before s:1sg.o:3sg see little
‘This child has grown since I have saw him last time.’ Tikuna (prim. data)
A.  i-pata-wa nucha dau [wii güena] tare buun nema cuchia-wa
   1sg-house-LOC s:1sg.o:3sg see one woman two child there kitchen-LOC
   ‘When I came home, I saw a women with two children in the kitchen.’

B: nukü wa [*nema / (lliema) güena]?
   s:2sg.o:3sg know DEM / 3SG.F woman
   ‘Did/do you know this woman?’

Tikuna (prim. data)

Instead of the demonstrative, an adnominal personal pronoun can be used with the anaphoric referent in (3). At this point, it is not clear how systematic the use of the pronoun as an adnominal anaphoric marker in Tikuna is. However, it is plausible that anaphoric articles in general develop from constructions like the one shown in (3). The demonstrative in Limbum shows a similar behaviour to its counterpart in Tikuna; the two languages differ in that Limbum additionally has an anaphoric article. Example (4) below shows the Limbum distal demonstrative čà in a deictic contexts. An example of an anaphoric referent is given in (5); it requires the use of the anaphoric article fɔ (cf. section 4.2.1).

(4) wɛ̀ riŋ [ndjiŋwɛ̀ čà/ *fs] ā?
   2sg know woman DEM:DIST ART:ANA Q
   ‘Do you know that woman (over there)?’
   Limbum (prim. data)

(5) A: mú mfɛʔ mɛ̀ bāʔ mɛ̀ yɛ̀ ndjiŋwɛ̀ bōō bāā ō čútèh mi
   when time 1sg PST2 arrive 1sg PST2 see woman and child.PL two 3PL sit in
   kitchen
   ‘When I came home, I saw a woman with two children in the kitchen.’

B: wɛ̀ mú riŋ [ndjiŋwɛ̀ fɔ/ *čà] ā?
   2sg PST2 know woman ART:ANA DEM:DIST Q
   ‘Did you know that woman?’
   Limbum (prim. data)

Figure 7.3 sketches a system of the Limbum type with a demonstrative marker that is used for deictic referents and an anaphoric article to encode anaphoric referents.

![Diagram](image_url)

Figure 7.3: Demonstrative with anaphoric article
In the simplified system shown in Figure 7.3, the demonstrative is mainly used to mark deictic referents, while the anaphoric article is used in contexts with anaphoric definite referents. Neither marker codes uniqueness-based definite referents systematically. This does not necessarily exclude the occurrence of a referential marker with situationally unique referents, but the latter do not require systematic encoding in this type of systems.

Demonstratives of type II which encode both deictic and anaphoric referents are crosslinguistically common. The following examples from Ejagham (Bantu, Cameroon) show such a demonstrative used for spatial deixis (6) and anaphora (7):

(6) ma ŋgole 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 (prim. data)

(7) A: ewugu m-ba ngi mɔninki na aβɔ aβi ka kifɔn 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 məŋe wi [mɔninki ɲu]?
    PST know.2sg cl1 woman.cl1 DEM:PROX.cl1
    ‘Did you know this woman?’

    Ejagham (prim. data)

The marker in Ejagham is a demonstrative and not an article, since it cannot mark situationally unique referents. Example (8) shows this for the referent of ndzuk oβasi ‘church’ which is situationally unique in the context of a village:

(8) [ndzuk oβasi (*ɲi)] 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 (prim. data)

Such a system with a demonstrative that encodes deictic and anaphoric but not situationally unique referents is illustrated in Figure 7.4 below.

---

2The assumption of demonstratives being generally unable to mark uniqueness-based definite referents is not unproblematic, since restrictive relative clauses are often marked by demonstratives on the head noun. I will not include this special use of demonstratives here, since it is a minor and formally restricted phenomenon.
Purely anaphoric markers are anaphoric articles (type III), like the anaphoric article in Limbum shown in example (5). Definite articles of type IV seem to be rather rare crosslinguistically. Such articles combine the functions of marking anaphoric and situationally unique referents, but are not used in deictic contexts. A candidate for this type of definite articles can be found in Indonesian. Indonesian uses the third person possessive marker -nyan to mark definite referents, also in contexts in which no possessor is semantically recoverable (Rubin 2010). Because of its distinct source, this marker was never used to encode deictic referents to begin with and has (not yet) extended to such contexts. Its use to mark situationally unique referents is shown in (9). In addition, we see in (10) that the marker -nyan can also encode anaphoric referents.

(9) kalau mau makan nasi-nya di lemari
If you want to eat rice-DEF in pantry
‘If you want to eat, the rice is in the pantry.’ Indonesian (Rubin 2010: 107)

(10) 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 deket rumah?
salon-DEF in near house
‘Is the salon near your house?’ Indonesian (Rubin 2010: 109)

Figure 7.5 shows a schematic representation of the Indonesian system. It consists of a demonstrative that primarily marks deixis, but that be used to encode anaphoric referents as well. In addition to this demonstrative, there is an article whose main function is the coding of situationally unique definite referents, also being able to mark anaphoric referents.

![Figure 7.5: Demonstrative with uniqueness-based definite article](image)

Definite articles of type V that can encode all three major types of definite referents probably correspond to the most common type of definite articles. A system with this type of definite articles typically uses the demonstrative to express deictic but also anaphoric referents. The definite article on the other hand encodes situationally unique referents, but it can also be used for the

---

3The situation that we find in many Uralic languages with a possessive marker is very similar in that these definiteness markers might also be definite articles of type IV.
other two referent types. In such a system, all three types of definite referents are systematically marked, as is shown in Figure 7.6.

These were the main functions that are relevant to distinguish between different types of definite articles and demonstratives. In addition to deixis, anaphora, and uniqueness, two other functions may be relevant for setting up the criteria to distinguish between demonstratives and definite articles. These two functions are the expression of distance and contrast.

Since demonstratives often encode the location of the referent with respect to the discourse situation, the marking of distance has been proposed to be a necessary criterion for demonstratives, especially regarding their distinction from articles (e.g. Anderson & Keenan 1985: 280). However, a closer look at demonstratives revealed that distance as such is not a necessary criterion. Especially with respect to adnominal demonstratives, we find languages with demonstratives that do not express distance (cf. Diessel 1999: 38, Diessel 2013a, Kemmerer 1999, Lyons 1999: 19f.). It seems that only adverbial demonstratives have at least a two-way spatial contrast ('here' vs. 'there'), and in many languages they can combine with an adnominal demonstrative in order to express distance (Diessel 1999: 36). The expression of distance or spatial deixis in the strict sense can therefore be considered to be a typical function of demonstratives in the world’s languages, but not a necessary one. Hence, the marking of distance cannot be used to distinguish between demonstratives and articles. Moreover, we find article systems that formally encode spatial deixis, which makes the latter even less appropriate as an exclusive property of demonstratives. A well-known example is the definite article in Macedonian which marks three levels of deixis or distance, as is shown in Table 7.1 (Lunt 1952: 41, Kramer & Mitkovska 2011: 211f.).

---

4The distribution of these three forms is not only conditioned by deixis but also by additional pragmatic and stylistic factors.
Table 7.1: Macedonian definite article paradigm

<table>
<thead>
<tr>
<th>deixis</th>
<th>singular masculine</th>
<th>feminine</th>
<th>neuter</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>ot</td>
<td>ta</td>
<td>to</td>
<td>te</td>
</tr>
<tr>
<td>proximal</td>
<td>ov</td>
<td>va</td>
<td>vo</td>
<td>ve</td>
</tr>
<tr>
<td>distal</td>
<td>on</td>
<td>na</td>
<td>no</td>
<td>ne</td>
</tr>
</tbody>
</table>

Another criterion used to define demonstratives is the expression of contrast. For instance, Hawkins (1978: 156f.) notes that demonstratives always involve a contrast between two referents, either explicitly or implicitly. He implements the notion of contrast as a matching constraint that applies to demonstratives, i.e. the referent is matched with another explicit or implicit referent, from which the former has to be disambiguated. According to Hawkins (1978: 157), definite articles automatically exclude contrast due to their uniqueness component. Similarly, Diessel (1999: 2) notes that “all languages have at least two demonstratives that are deictically contrastive […],” although this does not necessarily apply to adnominal demonstratives. A well-known example comes from German, which has the two contrastive adnominal demonstrative forms *dies* and *das*. Functionally, however, the markers do not distinguish between different distance values. Only in combination with the adverbial demonstratives, a distance contrast can be expressed, as is shown for *das* and *dies* in (11) and (12), respectively:

(11) a. *das* hier

    that here

    ‘this one’

    b. *das* da

    that there

    ‘that one’

(12) a. *dies* hier

    this here

    ‘this one’

    b. *dies* da

    this there

    ‘that one’

    German (Himmelmann 1997: 53)

Therefore, Himmelmann (1997: 53-62) argues against contrast as a necessary criterion for demonstratives. While contrast in terms of spatial deixis is arguably not a criterion for demonstratives, I follow Hawkins (1978) in that demonstratives typically have a contrastive component, which distinguishes them from definite articles. To the best of my knowledge, contrast in connection with definite articles has only been proposed for German. Schwarz (2009: 34) provides the following example showing that the definite masculine article *der* is used to express contrastive meaning:
The capitalization of the definite articles in (13) indicates that the articles bear emphatic stress, while definite articles in German are normally unstressed. This suggests that the elements in (13) are not regular articles but other elements that may qualify as demonstratives based on their spatial deictic and contrastive functions. Although treating them as articles, also Schwarz (2009: 34) notes: “[...] the strong article has demonstrative uses in addition to the anaphoric ones. However, such uses typically involve a pitch accent on the determiner, which suggests that they have a special status” (Schwarz 2009: 34). Based on this stress difference between regular definite articles and these article-like elements in German, I do not regard the latter as articles. This is in accordance with the crosslinguistic tendency of definite articles against the encoding of contrast, which in turn shows that the marking of contrast may serve as an additional criterion to distinguish between demonstratives and definite articles.

7.2 The numeral ‘one’

Not too differently from the situation seen for demonstratives and definite articles, the proper distinction between indefinite or exclusive-specific articles and the numeral ‘one’ is difficult to make in many cases, which often leads to confusion about the ontological status of articles. Do they correspond to a category, a function that certain markers, i.e. the numeral ‘one’, can have, or are they markers themselves? In this study, I use the notion of articles in two ways: articles correspond to a crosslinguistic category on a language-independent level, and on a language-specific level, they correspond to linguistic markers. I do not view articles as functions.

That this is still an open issue can be seen by the different statuses that are ascribed to articles, which partially results from the fact that articles are so similar to other elements. An example is the following description from Romero-Méndez (2008: 270) on Ayutla Mixe (Mixe-Zoque, Mexico): “The numeral tu’uk ‘one’ is commonly used as an indefinite article.” This description is then illustrated with the following two examples, both introducing a not yet identifiable but discourse-prominent referent:

---

For a more detailed discussion, see e.g. Engel (1988: 535,660); Himmelmann (1997: 50-56); Bisle-Müller (1991: 62-66); Hoffmann (2009: 312f.).
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 specific or indefinite article.

The distinction between indefinite articles and the numeral ‘one’ is problematic because of 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). Heine (1995: 71-76) suggests the following stages with regard to the development of the numeral and the indefinite article:

<table>
<thead>
<tr>
<th>Stage 0</th>
<th>Stage I</th>
<th>Stage II</th>
<th>Stage III</th>
<th>Stage IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeral</td>
<td>Presentative marker</td>
<td>Specific marker</td>
<td>Nonspecific marker</td>
<td>Generalized marker</td>
</tr>
</tbody>
</table>

Figure 7.7: Grammaticalization path from the numeral ‘one’ to a nominal marker

The stages from the scale in Figure 7.7 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 referents into the discourse. 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)
Similarly to the distinction between demonstratives and different types of definite article addressed in section 7.1, I argue that the distinction between the numeral ‘one’ and the article can be made based on a cut-off point on the scale presented in Figure 7.7, using it as a synchronic hierarchy of related functions (and not necessarily as an indicator of a grammaticalization path). This approach is innovative insofar, as 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 7.2 contains the properties provided in Heine (1997) based on Heine et al. (1995) that make reference to the numeral ‘one’ but cannot be used to distinguish between the latter and an indefinite article.

| Table 7.2: 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. |

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 due to grammaticalization usually correlates with the degree of semantic changes and extensions, the form development is only an epiphenomenon to the function development and cannot be a criterion as such. The necessary criteria should be based on the functions of the markers instead.

The different functions of the four stages including the numeral ‘one’ and different types of indefiniteness markers are shown in Figure 7.8. Again, we can distinguish between different markers based on the combination of functions they encode.
The stage 0 marker, 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 1978a: 256; Dixon 1980: 107-108; Heine 1997: 24).6

For the presentative function of stage I, i.e. the introduction of new 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’. Crucially, this use of the numeral one is confined to referents which are prominent and salient in the discourse. A specific but not discourse-prominent referent is not included in the function of stage I.

One example of a marker that encodes this presentative function is found in Spoken Hebrew. Givón (1981: 36) notes that a phonetically reduced form -xad of the numeral exad ‘one’ can mark a referent that is discourse-prominent as nonidentifiable.7 In (16), ish ‘man’ is the subject and topic of the sentence and its referent is encoded as nonidentifiable to the hearer by the marker -xad. In example (17), on the other hand, ish ‘man’ is contrasted with isha ‘woman’, which makes it a focused constituent. In this case, ish ‘man’ occurs as a bare noun.

---

6The famous counter-example to this claim is Pirahã (Mura, Brazil). The two numeral markers hói and hoí were first analyzed 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’.

7Note that Heine (1997: 72f) also refers to this example from Givón (1984: 36), but discusses it as a specific marker of stage II. I treat it as a presentational marker, since Givón shows how this marker is restricted to topical specific referents, or at least that it is not used systematically in other specific indefinite contexts.
Another example to illustrate this presentative use of the numeral ‘one’ comes from Tikuna (Tikuna-Yuri, Colombia) in (18). Although the use of the numeral in such contexts is possible in Tikuna, it appears to be optional or conditioned by other factors, as we see in examples (19) and (20).

(18) nücha dau [wii] hipata hichi llea 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 (prim. data)

(19) i-pata-wa nucha dau [(wii)] güena tare buun nema 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 [wii] a woman with two children in the kitchen.’ Tikuna (prim. data)

(20) nukü dau [(wii)] ail!
    s:2sg.o:3sg see one.pres tiger
   ‘Look! A tiger!’ Tikuna (prim. data)

Because the numeral ‘one’ is widely attested in this function, I view the presentative function 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’, I treat this marker as a presentational article (which in turn is a subtype of indefinite articles). I treat presentational articles differently from the numeral ‘one’ because the systematic coding of such referents exceeds the presentative function that is generally available for the numeral ‘one’ (cf. section 5.3.4).

Markers of type II correspond to exclusive-specific articles and 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 (21) and (22) below show the article ąa(k) with a discourse prominent referent; in (23), the noun baṭ-ā ‘stone’ occurs with the exclusive-specific article although its referent is less prominent (cf. section 5.1).
Another type of articles that can be distinguished based on the scale in Figure 7.8 is the indefinite article, which encodes specific and nonspecific referents. The marking of both referents and the resulting ambiguity was shown for English in example (21) which is repeated here as (24). For more examples of indefinite articles, see section 5.3.

(24)  

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 Figure 7.8 are nonspecific articles. These articles mark only nonspecific referents and do not occur with specific referents. An example of a nonspecific article in Q’anjobal is repeated in (25) below (cf. section 5.2).

(25)  

ay mi qjunoq yatut thioxh b’ay yich calle?  
exist Q ART:NSPEC house.of.God at back street  
‘Is there a church at the end of the road?’  
Q’anjobal (prim. data)

As we saw in this section, the numeral ‘one’ and different types of articles from the indefinite domain can be distinguished based on the functions and distributions that the markers have, assuming the relevant functions or semantic values of such markers to be individuation (quantification), the presentative function, specificity, and nonspecificity. Therefore, we do no longer depend on a catalogue of unrelated properties to define indefinite articles and delimit them from the numeral ‘one’. 
7.3 Negative polarity items

Negative polarity items (NPIs) are a heterogeneous group of elements such as any, ever, at all, either, and yet. What all these elements have in common and what motivates their label is that they are sensitive to polarity and that they typically occur in environments of negation. Some examples from English are given in (26) to (28).

(26)  a. Peter has not read any linguistics paper today.
     b. Peter has read *any linguistics paper today.

(27)  a. Peter does not want to read about semantics at all.
     b. Peter wants to read about semantics *at all.

(28)  a. I do not see anyone.
     b. I see *anyone.

The group of polarity-sensitive elements that is relevant for the purposes of this section can be narrowed down to the ones that occur in the nominal domain together with a noun, as does any in (26). Being sensitive to polarity does not restrict the occurrence of NPIs to contexts of negation. Since NPIs have been a popular topic in semantics for the last 40 years, I cannot mention and summarize all the aspects of NPIs that previous work has dealt with. In this section, I focus on the aspects that are important with regard to nonspecific articles. The relevant uses of NPIs cover a wide range of contexts (Havel 1997: 52; Givón 1984: 441-449; Giannakidou 2017), and are given in Table 7.3.

---

8Work on NPIs goes back to Baker (1970), Fauconnier (1975), Horn (1972), Ladusaw (1980) who discuss NPIs in English focusing on negation contexts. Amongst others, Giannakidou (1997) and 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).
Table 7.3: Relevant contexts for adnominal NPIs

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>negation</td>
<td>I didn’t see any cake in the fridge.</td>
</tr>
<tr>
<td>indirect</td>
<td>I don’t think there was any cake in the fridge. Tabea wrote a dissertation without having any coffee.</td>
</tr>
<tr>
<td>inherent negative predicates</td>
<td>Tomas failed to write any paper.</td>
</tr>
<tr>
<td>question</td>
<td>Did you drink any coffee today?</td>
</tr>
<tr>
<td>conditional</td>
<td>If you see any problems with my analysis, please let me know.</td>
</tr>
<tr>
<td>irrealis</td>
<td>TAM marking, modals</td>
</tr>
<tr>
<td></td>
<td>He could have bought any dog.</td>
</tr>
<tr>
<td>inherent irrealis predicates</td>
<td>She is looking for any student who can help her with the organization.</td>
</tr>
<tr>
<td>comparison</td>
<td>Cheetahs can run faster than any human.</td>
</tr>
<tr>
<td>free choice</td>
<td>You can take any cookie.</td>
</tr>
</tbody>
</table>

Without going into detail with respect to the semantic and pragmatic conditions of each type of contexts illustrated in Table 7.3, we can make use of the concept of veridicality proposed and argued for in 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 is one where truth inference seems to be suspended”. She provides the following example for illustration:

(29) a. Unfortunately, Mary saw a snake. ⇒ Mary saw a snake.
    b. Maybe Mary saw a snake. ⊳ Mary saw a snake.
    (Giannakidou 2017: 21)

In example (29a), 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 (29b), on the other hand, lies within the scope of the modal maybe, which blocks the inference of the truth value of the proposition expressed by that sentence. This is an example of a nonveridical statement.

---

9 I will not be concerned with two uses that are usually attributed to NPIs: free choice contexts as you can take any cookie and comparisons as in cheetahs can run faster than any human. The reason for excluding such contexts is that even though they are often expressed by NPIs, we find languages that do not use NPIs, but may have a separate marker. I take these contexts to be functional extensions and not a core function of NPIs.
Hence, nonveridicality can be taken as the property of propositions that licenses NPIs. Negation is a special case, as it does not only involve nonveridicality (the truth value of \( p \) cannot be assessed) but antiveridicality, which asserts that the proposition does not hold (\( p \) is not). In addition, Giannakidou (2017: 7) mentions an antiepisodicity constraint: if the proposition makes reference to a single past event, even e.g. negation cannot license NPIs as, e.g., free choice items:

\[
\begin{align*}
(30) & \quad \text{a. Any cat hunts mice.} \\
& \quad \text{b. \textbf{?}Any cat hunted mice.}
\end{align*}
\]

In (30a), \textit{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 proposition type allows for reference to multiple nonspecific events. In (30b), on the other hand, the proposition is tied to a single past event, which blocks the nonspecific free choice interpretation of \textit{any} and makes the sentence infelicitous.

The contexts for nonspecific articles are exactly the same, as could be seen in section 5.2. 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 \textit{any} in English and nonspecific articles lies in their distribution. Languages such as English, where the indefinite article \textit{a}, depending on the context, leaves the referent ambiguous between a specific and nonspecific interpretation, do not need to resolve the ambiguity between a specific and nonspecific referent. For this reason, we can use the indefinite article \textit{a} or a bare noun instead of the NPI \textit{any} in all contexts that were presented to be relevant to NPIs in Table 7.3:

\[
\begin{align*}
(31) & \quad \text{a. I did not see \textit{cake} in the fridge.} & \text{(negation)} \\
& \quad \text{b. I don’t think there was \textit{cake} in the fridge.} & \text{(neg-raising)} \\
& \quad \text{c. Tabea wrote a dissertation without having \textit{coffee}.} & \text{(indirect negation)} \\
& \quad \text{d. Tomas failed to write \textit{a paper}.} & \text{(negative verb)} \\
& \quad \text{e. Did you drink \textit{coffee} today?} & \text{(question)} \\
& \quad \text{f. If you see a \textbf{problem} with my analysis, please let me know.} & \text{(conditional)} \\
& \quad \text{g. He could have bought \textit{a dog}.} & \text{(irrealis)} \\
& \quad \text{h. She is looking for a \textbf{student} who can help her with the organization.} & \text{(irrealis verb)}
\end{align*}
\]

Of course, even though the utterances from Table 7.3 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 7.3 and in example (31) above. We find differences of two types. Either, using the article or a bare noun, the referent is ambigu-
ous between a specific and nonspecific interpretation in the contexts in which no other explicit linguistic clue for nonspecificity is provided. We see this in (31g) and (31h). Or, NPIs can involve pragmatic effects that are absent in the sentences in (31). 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 section 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, we could see for English that NPIs are not necessary in the sense that there are linguistically simpler alternatives (bare nouns, the indefinite article) that can have the same semantic interpretation.

To show that we see a similar effect in languages without articles, the following Polish examples (32) to (37) show that we can equally find corresponding expressions of nonspecific referents either marked by an NPI (żaden, jakiś, or jakikolwiek) or left as a bare noun:

(32) Nie widziałam w lodówce żadnego ciasta.
'I didn’t see (any) cake in the fridge.’ Polish (prim. data)

(33) Tabea powiedziała to bez żadnej ironii.
‘Tabea said it without (any) irony.’ Polish (prim. data)

(34) Tomas odmówił napisania jakiegokolwiek artykułu.
‘Tomas declined to write (any) articles.’ Polish (prim. data)

(35) Czy piła dzisiaj kawę?
‘Did you have coffee today?’ Polish (prim. data)

(36) Daj mi znać, jeśli zauważyasz jakieś błędy w mojej analizie.
‘Let me know, if you notice (any) errors in my analysis.’ Polish (prim. data)

(37) Szuka jakiegoś studenta, który mógłby jej pomóc z organizacją.
‘She is looking for a (any) student who can help her with the organization.’ Polish (prim. data)

Being beyond the scope of the present study, I do not address the functions and distributions of different NPIs and free choice items in Polish here. For a discussion of NPIs in Polish, cf. Błaszczyk (1999, 2003).

I thank Joanna Zaleska for providing these Polish examples.

The referent that belongs to the word artykuł ‘article’ is ambiguous between a definite and nonspecific interpretation without the NPI jakiegokolwiek and further context.

In polar questions, NPIs do not seem acceptable in Polish.
After this digression on NPIs and their contexts, we return to the initial question of this section, i.e. the difference between NPIs and nonspecific articles. As noted earlier, the semantics (referential functions) of NPIs do not differ from the ones of nonspecific articles. Their difference rather lies in the additional pragmatic effects that NPIs have and nonspecific articles lack.

The preceding paragraphs showed that although *any* in English is a dedicated marker of nonspecificity because it occurs in various nonspecific contexts, it is not systematically required to resolve the referential vagueness between specific and nonspecific referents. Generally, bare nouns or nouns with an indefinite article can also be used in nonspecific contexts without resolving the ambiguity. In languages with no indefinite article, referents expressed by bare nouns can also receive a nonspecific interpretation, which was exemplified by the Polish examples in (32) to (37). However, NPIs like *any* in English or *żaden*, *jakiś*, and *jakikolwiek* in Polish are not necessarily required, since a simpler expression is available.

On the other hand, nonspecific articles occur systematically to resolve this ambiguity between a specific and a nonspecific referent. As was already shown in section 2.2.2, Q’anjobal has a nonspecific article. Example (38a) illustrates that the specific indefinite article *jun* is not acceptable in nonspecific contexts. Its nonspecific counterpart *junoq*, on the other hand, is obligatorily used in contexts in which the referent is interpreted as nonspecific. Moreover, (38b) shows that the use of a bare noun is not allowed either. This suggests that, indeed, *jun* is a specific and not an indefinite article, and that *junoq* is not an adnominal NPI, but a nonspecific article, being systematically required to mark the referential value of nonspecificity.

(38) a. asi’ go.IMP [junoq/ *jun*/ tuktuk].
   yul ART:NSPEC ART:SPEC mototaxi
   ‘Take a/any tuktuk.’
   Q’anjobal (prim. data)
   b. *asi’ go.IMP [tuktuk].
   yul ART:SPEC mototaxi
   ‘Take a/any tuktuk.’
   Q’anjobal (prim. data)

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 them as nonspecific articles.

### 7.4 Summary

In this chapter, I pointed out how certain types of articles can be distinguished from similar elements. The distinction between definite articles and demonstratives was shown to be problem-
atic because of their diachronic connection. I argued that deictic markers which can also mark anaphora are not articles, but should be treated as demonstratives. Anaphoric markers that cannot be used to encode deixis, on the other hand, are treated as anaphoric articles. The third main type of articles in the definite domain are definite articles, marking anaphora, situational uniqueness, and often deixis as well.

Mutatis mutandis, I argued that numerals which can extend to presentative uses should not be considered as articles, because this functional extension of the numeral ‘one’ seems to be generally available across languages. If discourse-prominent and nonidentifiable referents are systematically encoded, however, I treat such markers as presentational articles, also if they formally correspond to the numeral ‘one’. 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 referents as nonspecific in addition, it corresponds to an indefinite article.

The last part discussed the distinction between nonspecific articles and negative polarity items (NPIs). Negative polarity items are typically licensed under the scope of polarity, but other main contexts for NPIs include questions, conditionals, and irrealis contexts. Adnominal NPIs have similar functions to the markers that I treat as nonspecific articles. I argued that the difference between nonspecific articles and NPIs thus does not lie in their referential functions, but in the systematicity of their distribution in nonspecific contexts. Nonspecific articles occur systematically to encode a referent as nonspecific, while NPIs are not systematically required in such contexts, and if they occur, they often have additional pragmatic effects that nonspecific articles lack.
Chapter 8

Crosslinguistic trends and variation

In this chapter, I discuss a number of trends and patterns of variation of articles in the world’s languages, starting with the morphological properties of articles in sections 8.1 and 8.2. Section 8.1 discusses the inflection of articles by first giving an overview of the crosslinguistic patterns of article inflection. Then, the two main factors that can account for the inflectional behavior of articles are discussed: the inflection of the noun in the presence of the article, as well as the inflectional properties of the article’s source element. Section 8.2 addresses a number of questions concerning article paradigms and presents the complex article paradigm of Mokpe in detail. Section 8.3 is concerned with the crosslinguistic distributions of different article types. The first part presents a number of areal trends, the second part relates crosslinguistic distribution of article types to the expression of referent types, and the last part discusses two unattested article types. The last part of this chapter, section 8.4, deals with crosslinguistic trends concerning article systems. I first propose a typology of article systems on the basis of the referential space covered by the system, and then present the crosslinguistic trends regarding article systems in terms of that typology. Despite a high degree of variation, it can be shown that the frequencies of attested systems and also the gaps can be accounted for by two general principles. Systems with functionally overlapping articles are crosslinguistically rare, but attested. I present a few of those systems in the last part of this section, showing that they also follow a more general pattern.

8.1 Article inflection

This section addressed the inflectional properties of articles. I provide a number of examples to illustrate and discuss the inflectional patterns of articles, and show that we find a crosslinguistic trend against the inflection of articles. While a number of articles especially in European languages may suggest the contrary, the impression that articles often do not inflect is not a new one and already found as an impressionistic observation in Lyons (1999: 1967):
The definite article is, in most languages having one, an invariable word, in the sense that it is not inflected (though it may undergo some allomorphic variation, as with Hungarian az pre-vocally and a pre-consonantally). Thus [...] English the show[s] no agreement with the head noun or other constituents of the noun phrase. This is often the case even in languages with a fairly high degree of inflection, in which a related demonstrative does show inflection. Thus in Hungarian, the demonstrative az takes case inflections as does the head noun, but the article az, a, derived historically from it, does not.

However, this has not yet been systematically tested across languages. I show in this section for number, gender, and case marking that we find two main factors which restrict their marking on the article: the expression of number and case on the noun as well as the inflectional properties of the source elements of articles.

8.1.1 Selected patterns of article inflection

In the world’s languages, the categories that are marked on the article include number, gender, case, and deixis. In this section, I present examples of articles that inflect for different categories, focusing on the marking of gender, case, and deixis. Since section 8.1.3.1 deals with the association between article and noun inflection, the discussion of article inflection here will focus on its interaction with nominal inflection as well.

Gender marking on the article is common across languages, in fact, gender is the nominal category with the strongest trend to be marked on the article (cf. section 8.1.2). Before we can address the question of gender marking on articles, a few preliminary remarks about gender are required.

I take a language to have gender if we can distinguish different groups of nouns based on their behaviour with respect to agreement on other elements, e.g., verbs, adjectives, demonstratives, quantifiers, numerals, or the article (cf. Corbett 2003, Corbett & Fedden 2016). I refer to those different groups as classes of a gender system. I do not distinguish between different types of gender systems; purely sex-based or animacy-based systems are included if they fulfill the agreement criterion, as well as “noun class” systems and systems with nominal classifiers. I do not further distinguish between these types of gender systems. This is because a more detailed and accurate distinction is hardly possible without a lexical analysis for each language, which exceeds the purposes of this study. In other words, I include languages with a distinction between masculine and feminine “genders” for human referents (e.g. Dime, Konso, Wayuu, Asheninka Perené), “sortal classifier” systems (e.g. Cubeo, Tepehua, Ch’ol, Rajbanshi), “noun class” systems similar to the “Bantu type” (e.g. Runyankore, Bemba, Mokpe), and “gender” systems found in e.g. many Indo-European and Australian languages (e.g. Domari, Palula, Guragone, Wardaman). The same holds for animacy, treated as gender if animate and inanimate nouns show differences in
agreement. In the sample, only a small number of languages have animacy-based gender (e.g. Ute and Biak).

Important is also the article as the target for gender marking: I count an article as inflecting for gender if it has a different exponent for at least one of the gender classes in the language. Since I base gender on agreement between different classes of nouns and other elements, I categorically exclude gender marking on the noun. This is not to say that nouns do not contain information on the gender class in their phonological form. A case in point is Spanish.¹ For Spanish, we can establish phonology-based rules for gender class assignment that cover a large portion of the nouns in the language. The following two rules can assign a gender class to the majority of Spanish nouns:

(1) a. -a → feminine 
b. -o → masculine

Even though this rule can be applied successfully in most cases, there are two reasons against treating Spanish nouns as gender-marking, following the argumentation of Harris (1991). Firstly, there are forms for which the rules predict the wrong gender class, e.g., the noun problema ‘problem (m)’ or the noun mano ‘hand (f)’. Secondly, there are many nouns in the language that do not have either the final vowel -a/-o or any suffix that could automatically assign the gender class, which means that the rules shown in (1) are not applicable to a number of nouns in the language. Some examples are sol ‘sun (m)’, col ‘cabbage (f)’, tribu ‘tribe (f)’, espíritu ‘spirit (m)’.

A well-known example of an article that is inflected for gender is found in German. As Table 8.1 shows, both the definite and the indefinite article have different exponents according to the gender class of the noun that they occur with. However, the definite article does not inflect for gender in the plural. Since this is systematic in that it can be captured as a syncretism between the three gender classes, it is uncontroversial to regard the articles in German as gender-marking.

<table>
<thead>
<tr>
<th>masculine</th>
<th>singular feminine</th>
<th>neuter</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>def</td>
<td>indef</td>
<td>def</td>
</tr>
<tr>
<td>NOM</td>
<td>der</td>
<td>ein</td>
<td>die</td>
</tr>
<tr>
<td>ACC</td>
<td>den</td>
<td>einen</td>
<td>die</td>
</tr>
<tr>
<td>DAT</td>
<td>dem</td>
<td>einem</td>
<td>der</td>
</tr>
<tr>
<td>GEN</td>
<td>des</td>
<td>eines</td>
<td>der</td>
</tr>
</tbody>
</table>

¹For a similar discussion of gender marking on nouns in German, cf. Köpcke (e.g. 1982), Köpcke & Zubin (1984).
A perhaps less intuitive example for gender marking on the articles comes from Ch’ol (Mayan, Mexico). Ch’ol uses numeral classifiers for nouns that are used together with numerals. The classifiers, occurring as suffixes on the numerals, are presented in Table 8.2 (Vázquez Alvarez 2011: 160).

<table>
<thead>
<tr>
<th>classifier</th>
<th>semantic properties of the noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>-p’e(j)</td>
<td>/ (default classifier)</td>
</tr>
<tr>
<td>-tyikil</td>
<td>person</td>
</tr>
<tr>
<td>-kojty</td>
<td>animal, chili</td>
</tr>
<tr>
<td>-tsijty</td>
<td>small, thin, large things</td>
</tr>
<tr>
<td>-k’ej</td>
<td>flexible, square rectangle or rounded things</td>
</tr>
<tr>
<td>-sejl</td>
<td>rounded things</td>
</tr>
<tr>
<td>-lejch</td>
<td>flat things</td>
</tr>
<tr>
<td>-tyejk</td>
<td>tree</td>
</tr>
</tbody>
</table>

The classifier -p’e(j) is the default classifier used for the majority of the nouns in Ch’ol. The indefinite article in Ch’ol originates from the numeral ‘one’; therefore, it obligatorily contains one of these classifier suffixes according to the noun it occurs with. Based on the data provided in Vázquez Alvarez (2011), the indefinite article does not occur without being marked by a classifier. Since the choice of classifier depends on the noun, I regard the presence of the classifier on the indefinite article as gender marking for the purposes of the present study. Examples for the gender marking on the indefinite article are given in (2) to (5) below:

(2) ta’=bi i-tyaj-a [jum-p’e juñ] PFV=QUOT 3SG-find-TR ART:INDEF-CL:PEJ paper ‘It is said that he found a document.’ Ch’ol (Vázquez Alvarez 2011: 92)

(3) poj añ lonk-sa’ tyi [jum-p’e bolsaj] HON EXIST 1PL.EX-pozol in ART:INDEF-CL:PEJ plastic.bag ‘We have our pozol in a plastic bag.’ Ch’ol (Vázquez Alvarez 2011: 171)


(5) añ-bi jiñi [juñ-tyikil wiñik] EXIST-QUOT HESIT ART:INDEF-CL:TYIKIL man ‘There, hm, was a man.’ Ch’ol (Vázquez Alvarez 2011: 245)

As was mentioned in section 6.1.1, the inclusive-specific article in Bemba is prefixed to the noun, preceding the noun class marker that is traditionally regarded as an indicator of number and
gender. Table 8.3 below shows the article exponents and the noun class markers again, rearranged into the gender classes A to H split up into the singular and plural (the labels of the classes are chosen arbitrarily for convenience).

<table>
<thead>
<tr>
<th>gender class</th>
<th>article</th>
<th>noun class prefix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>singular</td>
<td>plural</td>
</tr>
<tr>
<td></td>
<td>singular</td>
<td>plural</td>
</tr>
<tr>
<td>A</td>
<td>u-</td>
<td>a-</td>
</tr>
<tr>
<td>B</td>
<td>u-</td>
<td>i-</td>
</tr>
<tr>
<td>C</td>
<td>(i-)</td>
<td>a-</td>
</tr>
<tr>
<td>D</td>
<td>i-</td>
<td>i-</td>
</tr>
<tr>
<td>E</td>
<td>i-</td>
<td>-N-</td>
</tr>
<tr>
<td>F</td>
<td>u-</td>
<td>a-</td>
</tr>
<tr>
<td>G</td>
<td>u-</td>
<td>u-</td>
</tr>
<tr>
<td>H</td>
<td>u-</td>
<td>-ku-</td>
</tr>
</tbody>
</table>

Comparing the noun class markers and the article exponents, we see that the article exponent always corresponds to a copy of the vowel in the noun class marker. Therefore, the article exponent may rather be a result of a phonological process, which is why I do not treat the inclusive-specific article in Bemba as a gender-marking.

Another interesting case with respect to gender marking on the article and the noun is Supyire (Gur, Mali). Supyire has a definite article and, according to Carlson (1994), it has the five gender classes 1-5, shown in Table 8.4. Gender as a category of the noun in Supyire is manifested by agreement on adjectives and other elements in the noun phrase (Carlson 1994: 75). Nouns of class 1-3 have singular and plural forms, whereas the classes 4 and 5 contain mass nouns which do not exhibit a number distinction. Carlson (1994: 77) presents two paradigms for the gender marking of nouns in the language. One corresponds to what he calls “basic noun gender suffixes”, the other paradigm contains the “definite noun gender suffixes”. I treat the first forms as the forms of the bare noun, while the other forms correspond to the combination of the noun and the definite article. The abstract exponents for the definite article in Table 8.4 suggest that the article inflects for gender, given nouns of the different gender classes are shown to condition a different exponent of the article.
Table 8.4: Gender marking on Supyire nouns and articles

<table>
<thead>
<tr>
<th>gender</th>
<th>nominal marker</th>
<th>definite article</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>singular</td>
<td>plural</td>
</tr>
<tr>
<td>1</td>
<td>-wV</td>
<td>-(bi)li</td>
</tr>
<tr>
<td>2</td>
<td>-gV</td>
<td>-yV</td>
</tr>
<tr>
<td>3</td>
<td>-lV</td>
<td>-gili</td>
</tr>
<tr>
<td>4</td>
<td>-rV</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-mV/-bV</td>
<td></td>
</tr>
</tbody>
</table>

Examples of bare nouns and nouns with the definite article are presented in Table 8.5 (based on Carlson (1994: 79-94)). While the correspondence between the gender class and the nominal suffix is irregular, the article exponent is consistently conditioned by the gender class of the noun. The variation in the article exponents within classes that we see for the singular of class 2 and for class 4 (as well as the tonal effects) can be accounted for by regular phonological processes in the language. Therefore, I regard the definite article in Supyire as inflecting for gender. The Supyire patterns contrasts with the one in Bemba in an interesting way: it shows that the ability of the article to inflect does not depend on how close it is to the noun, given that the article is affixed to the noun in both languages.
Table 8.5: Examples of gender marking on Supyire articles

<table>
<thead>
<tr>
<th>gender</th>
<th>noun</th>
<th>noun+article</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>SG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>foo</td>
<td>foòŋi</td>
</tr>
<tr>
<td></td>
<td>nu</td>
<td>nûŋi</td>
</tr>
<tr>
<td></td>
<td>sò</td>
<td>sòŋi</td>
</tr>
<tr>
<td></td>
<td>PL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>zòmii</td>
<td>zòmpíi</td>
</tr>
<tr>
<td></td>
<td>dufáa</td>
<td>dufáabíi</td>
</tr>
<tr>
<td></td>
<td>pwùnn</td>
<td>pwùunbíi</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>űkéŋɛ̀</td>
<td>űkêŋke</td>
</tr>
<tr>
<td></td>
<td>canŋa</td>
<td>cànŋke</td>
</tr>
<tr>
<td></td>
<td>baga</td>
<td>bage</td>
</tr>
<tr>
<td></td>
<td>kafeɛge</td>
<td>kafeɛge</td>
</tr>
<tr>
<td></td>
<td>bâhà</td>
<td>bâhe</td>
</tr>
<tr>
<td></td>
<td>lwoŋɔ</td>
<td>lwohë</td>
</tr>
<tr>
<td></td>
<td>PL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ciye</td>
<td>ciyì</td>
</tr>
<tr>
<td></td>
<td>kùùyò</td>
<td>kùùyi</td>
</tr>
<tr>
<td></td>
<td>canya</td>
<td>canyi</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sàhàlà</td>
<td>sàhâni</td>
</tr>
<tr>
<td></td>
<td>cere</td>
<td>cení</td>
</tr>
<tr>
<td></td>
<td>fwoùu</td>
<td>fwuùni</td>
</tr>
<tr>
<td></td>
<td>PL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>shàhàŋii</td>
<td>shàhâŋkíí</td>
</tr>
<tr>
<td></td>
<td>jàhìi</td>
<td>jàhií</td>
</tr>
<tr>
<td></td>
<td>tugugii</td>
<td>tugugíí</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>suro</td>
<td>sure</td>
</tr>
<tr>
<td></td>
<td>shire</td>
<td>shiré</td>
</tr>
<tr>
<td></td>
<td>sɛɛɛɛ</td>
<td>sɛɛɛɛɛ</td>
</tr>
<tr>
<td></td>
<td>nana</td>
<td>nante</td>
</tr>
<tr>
<td></td>
<td>kòònɔ̀</td>
<td>kòônte</td>
</tr>
<tr>
<td></td>
<td>tɔɔnnɔ</td>
<td>tɔɔnnnte</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>sînŋkanma</td>
<td>sînŋkanmpé</td>
</tr>
<tr>
<td></td>
<td>bèŋnmè</td>
<td>bèŋnmpe</td>
</tr>
<tr>
<td></td>
<td>suùmè</td>
<td>suùmpè</td>
</tr>
</tbody>
</table>

For the purposes of the present study, animacy is also regarded as gender if it is reflected in agreement. However, animacy of the noun is rarely marked on the article in the world’s languages and is attested only in a small number of languages in the sample. One example for articles with different exponents marking the animacy of the referent comes from Ute (Numic, USA). Animacy can be regarded as a gender category in Ute, since it is reflected in the agreement on e.g. demonstratives and adjectives (Givón 2011: 50,60). The Ute article system consists of a definite
article which distinguishes between animate and inanimate referents in addition to number and case (subject vs. non-subject) of the noun. The paradigm, based on Givón (2011: 163), is given in Table 8.6:

<table>
<thead>
<tr>
<th></th>
<th>animate</th>
<th>inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>subject</td>
<td>non-subject</td>
</tr>
<tr>
<td>singular</td>
<td>‘u</td>
<td>‘uway</td>
</tr>
<tr>
<td>plural</td>
<td>‘umu</td>
<td>‘umu</td>
</tr>
</tbody>
</table>

Examples to illustrate the animacy distinctions are given in (6) to (8) below, contrasting the animate (6) and inanimate (7) singular subject markers, and the animate (8) and inanimate (6) object markers.

(6)   kh-‘ura [‘áapachi ‘u] [tukuavi ‘uru] tuka-püga
then-be boy.AN ART:DEF.AN meat.INAN ART:DEF.INAN eat-REMPST
‘then the boy ate the meat’ Ute (Givón 2011: 163)

(7)   [tuka’napu ‘uru] yaqh-kya
table.INAN ART:DEF.INAN break-ANT
‘the table broke’ Ute (Givón 2011: 164)

(8)   [tuachi ‘uway] [tukuavi ‘uru] magha-püga
child.AN ART:DEF.AN meat.INAN ART:DEF.INAN feed-REMPST
‘she fed the meat to the child’ Ute (Givón 2011: 164)

Case marking on articles and other determiners may not be unexpected from the Indo-European perspective. However, in the world’s languages, case marking on articles is relatively rare and it is much more frequently marked on the noun only. As for case marking, I only consider core cases in order to keep case marking comparable across languages. I take a language to have case marking if it systematically marks nouns as core arguments of intransitive and transitive constructions. Thus, if at least one marker consistently occurs with a noun to mark it as the sole S argument of an intransitive clause, and/or as the A and/or P argument of a transitive clause, I treat it as a case marker. What may be called case marking but what I do not treat as such here is the marking of nouns as oblique arguments, instruments, benefactives, recipients, genitives/possessives, or spatial relations, etc. (e.g. in Guragone, Irish, Ao).

One example of core case marking on the article alone is the article in German. Even though German nouns are traditionally presented in case-number paradigms, the form of nouns generally does not distinguish between its function as S/A argument (nominative) and as P argument.
(accusative). Table 8.7 shows that what is partially distinguished are the genitive and dative forms of the noun, but not the nominative and accusative forms.

Table 8.7: Case marking on nouns and the definite article in German

<table>
<thead>
<tr>
<th></th>
<th>masculine</th>
<th>feminine</th>
<th>neuter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>singular</td>
<td>plural</td>
<td>singular</td>
</tr>
<tr>
<td>Nom</td>
<td>der Baum</td>
<td>die Bäume</td>
<td>die Wiese</td>
</tr>
<tr>
<td>Acc</td>
<td>den Baum</td>
<td>die Bäume</td>
<td>die Wiese</td>
</tr>
<tr>
<td>Dat</td>
<td>dem Baum</td>
<td>den Bäumen</td>
<td>der Wiese</td>
</tr>
<tr>
<td>Gen</td>
<td>des Baums</td>
<td>der Bäume</td>
<td>der Wiese</td>
</tr>
</tbody>
</table>

The only marker on the noun that qualifies as a core case marker in German is -en, which marks accusative singular nouns of the weak declension type (Hentschel & Weydt 2003: 139). While it is still obligatorily used with some nouns, e.g. den Student-en ‘the acc student-acc’, other nouns that fall into this inflection class can occur without the marker, e.g. den Bär(-en) ‘the acc bear-acc’. Neither does the article distinguish case in all relevant contexts. As was shown in Table 8.1, the exponents of both the definite and the indefinite article are also highly syncretic. Only their masculine singular forms formally distinguish between the nominative and accusative case. Nevertheless, this distinction is systematic so that I consider both articles in German as case-marking.

Another example of case being marked on the article but not on the noun comes from Albanian. Usually, Albanian nouns are arranged into a definite and an indefinite paradigm which are marked for number and case. The so-called definite forms of the noun correspond to the combination of the noun and the definite article which occurs as a suffix on the noun. On the other hand, the so-called indefinite forms rather correspond to the default or citation form of the nouns. The definite article is suffixed to the noun and is fused with the case marker. Table 8.8 below illustrates the case markers for four different inflection classes, largely corresponding to masculine, feminine, neuter, and a plural nouns (based on Newmark et al. (1982: 159-162) and Buchholz & Fiedler (1987: 269-271)).

---

2 Buchholz & Fiedler (1987: 269) distinguish between 5 inflection classes; the case syncretisms for singular and plural nouns are essentially the same in all five classes.
As we see in Table 8.8, it is only the definite article with singular nouns that distinguishes between nominative and accusative forms. Nouns that occur without the definite article do not formally distinguish these two core cases across the four inflection classes. Therefore, I only treat the definite article but not the noun as marking case in Albanian.

As I will show in section 8.1.2, articles that mark case are not common across the languages of the world, and in the sample, paradigms similar to the ones in German or Albanian are in fact confined to a handful of languages in mostly Eurasia (German, Albanian, Basque) and Australia (Warrwa, Arrernte), and rarer but attested in North America (Ute), South America (Aguaruna), and Papunesia (Sundwadia).

A language that marks core cases both on the article and on the noun is Diyari (Karnic, Australia). In addition, the definite article in Diyari shows how different case syncretisms in nouns and articles can come about: the definite article originated from third person pronouns that distinguish between the nominative, accusative, and ergative, while the majority of nouns only distinguish between nominative and ergative cases. Table 8.9 shows the core case forms of the definite article in Diyari (Austin 2011: 68).

As we see in examples (9), (10), and (11) that both the article and the noun are inflected for case.

(9) \[\text{thana} \text{ wilha} \] \text{wima} \text{ kirli-yi} \\
\text{ART:DEF.F.PL.NOM} \text{woman.NOM} \text{corroboree.NOM} \text{dance-PRES} \\
The women are dancing a corroboree.' 

Diyari (Austin 2011: 123)
The two examples in (12) and (13) show the combination of an article with a nonfeminine noun that does not distinguish between nominative and accusative, while the article does.

(12) [nhawu kupa] muka thurara-yi
  ART:DEF.NF.NOM child.NOM sleep.ABS lie-PRES
  ‘The child is sleeping.’  
  Diyari (Austin 2011: 91)

(13) ngathu [nhinha kupa] wapa-lka-yi
  1SG.ERG ART:DEF.SG.NF.ACC child.NOM go-APPL-PRES
  ‘I take the child for a walk.’  
  Diyari (Austin 2011: 80)

Core case marking on both the article and the noun being rare in the world’s languages, the pattern in Diyari with its definite article that formally corresponds to the third person pronouns and retained its inflectional behaviour is an illustrative example of how such systems can develop, and it also shows that the case morphology of the article can remain independent of the case value of the noun.

Articles do not only inflect for nominal categories; it is not uncommon to find articles that mark some form of deixis, mostly in those cases in which the source element already marked deixis and the article retained the existing formal opposition. An example of a definite article that can mark deixis and that developed from a set of demonstrative markers can be found in Macedonian. Table 8.10 shows the article exponents with the demonstrative markers in brackets. The distance-neutral markers from the first row correspond to the default article exponents which are the least marked for deixis and also the most frequent forms in language use (Kramer & Mitkovska 2011: 212).

Table 8.10: Macedonian definite article paradigm

<table>
<thead>
<tr>
<th>deixis</th>
<th>masculine</th>
<th>singular</th>
<th>feminine</th>
<th>neuter</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>ot (toj)</td>
<td>ta (taa)</td>
<td>to (toa)</td>
<td>te (tie)</td>
<td></td>
</tr>
<tr>
<td>proximal</td>
<td>ov (ovoj)</td>
<td>va (ovaa)</td>
<td>vo (ova)</td>
<td>ve (ovie)</td>
<td></td>
</tr>
<tr>
<td>distal</td>
<td>on (onoj)</td>
<td>na (onaa)</td>
<td>no (ona)</td>
<td>ne (onie)</td>
<td></td>
</tr>
</tbody>
</table>

3Based on Lunt (1952: 41), Kramer & Mitkovska (2011: 211f).
This relation between three (neutral, proximal, distal) sets, the rows in Table 8.10, of article exponents that are not base on the “usual” gender, number, and case distinctions relates to a more general important issue. I treat these forms as exponents of a single definite article in Macedonian, additionally marking deixis. An alternative view would be to treat these three sets as three definite articles based on their deixis value, each being additionally marked for gender and number. However, since marking distance does not affect their referential function, in cases like the Macedonian definite article, deixis should be treated like number, gender, or case, corresponding to a category that is additionally marked on the article (cf. section 8.2.1).

The article exponents from the first row in Table 8.10 formally correspond to the medial demonstrative markers and represent the default definite article exponents. They can generally be used in all contexts that require definite articles. The proximal and distal forms are less frequent and mainly used in certain deictic contexts. However, they are not restricted to deictic functions and are conditioned by register, style, and creative language use according to my informant. Examples (14) and (15) illustrate the use of the proximal and distal forms of the definite article. As (14b) and (15b) show, the deixis neutral form of the article can equally be used in both cases.

(14) a. ima restoran vo.sel-vo?
   have.3sg restaurant in village-ART:DEF.PROX
   ‘Is there a restaurant in the / this village?’

   b. ima restoran vo.sel-to?
   have.3sg restaurant in village-ART:DEF
   ‘Is there a restaurant in the village?’

   Macedonian (prim. data)

(15) a. se sekavaš na ideja-na što ja imav?
   refl remember.pst.2sg on idea-ART:DEF.DIST comp I have.pst.1sg
   ‘Do you remember the / that idea that I had?’

   b. se sekavaš na ideja-ta što ja imav?
   refl remember.pst.2sg on idea-ART:DEF comp I have.pst.1sg
   ‘Do you remember the idea that I had?’

   Macedonian (prim. data)

Other cases of articles marked for deixis are the referential article in Halkomelem (cf. section 6.2.2) and the anaphoric articles in Bare. For the anaphoric articles, marking deixis corresponds to marking more or less time between the mentioning of the antecedent and the anaphoric referent. Examples (16) and (17) show this for Bare.

(16) tšinu yawi mehėsa me-wát’uka [da tšinu] i-báraka
    dog angry 3pl.want 3pl-beat ART:ANA.PROX dog 3pl.NF-run
    ‘The dog was angry, they wanted to beat the dog, it ran away.’ Bare (Aikhenvald 1995: 24)
8.1.2 Crosslinguistic trends

As was already mentioned in the previous section, articles inflect most frequently for gender, and less frequently for number and case. With respect to inflection in general, this section shows that we find a trend against the inflection of articles in the world’s languages.

What I treat as article inflection here is the systematic marking of number, gender, and case of the noun that the article occurs with. Gender and case marking is understood here as was defined in the previous section. I treat number marking accordingly. Number has to be marked systematically on the noun or the article in order to be counted as number marking. This includes systematic number marking for a subset of noun, e.g. only on nouns with human or animate referents (e.g. Oko, Tongan, Arawak, Wubuy). On the other hand, if number marking is described as “optional” in the nominal domain without further indication of the conditions for number (mostly plural) marking, I do not treat articles or nouns as number-marking (e.g. Siar Lak, Mangarrayi, Jingulu).

Based on the sample of 113 languages with 148 articles in total, Figure 8.1 shows the number of articles which inflect and which do not inflect, divided into the six macro areas as distinguished in Dryer & Haspelmath (2013), Hammarström et al. (2018). Upon visual inspection of the data in Figure 8.1, article inflection seems to depend on the area: Eurasia and North America show a large proportion of articles that do not inflect, Africa, South America, and Papunesia to a lesser extent, while we find slightly more articles in Australia that inflect than those that do not inflect.

Despite these differences, the overall trend shows more articles without inflection than articles that inflect. The sample of 148 articles contains 55 inflecting articles and 93 non-inflecting articles.

---

(17) nu-yadâ nu-tjâna-ka [asa tʃiŋu]  
1SG-see 1SG-stay-DECL ART:ANAL.DIST dog  
‘I am still seeing that dog.’

Bare (Aikhenvald 1995: 24)

---

4For details on the morphological properties of single languages and articles see Tables A.1 to A.6 in the appendix.
In order to test potential areal biases in more detail, a binomial regression model was fitted to estimate the underlying proportions of inflecting articles by macro area. The model was fitted using Bayesian methods (Markov Chain Monte Carlo sampling with Stan) with the “brms” package (Bürkner 2017, 2018) in R (R Core Team 2016). This also applies to all other fitted models in this chapter. The model fitted can be described as follows:

\[(N \text{ inflected} \mid \text{trials}(N \text{ total}) - \text{macro area})\]

In contrast to the “traditional” frequentist method, the Bayesian method does not assume that there is a single underlying coefficient for the proportion that is estimated whose probability is then evaluated against the null hypothesis. In a Bayesian model, the coefficient of the estimated proportion itself is modelled as a probability function: many different coefficients are possible, but they are not equally likely on the basis of prior assumption and the data observed. Thus, modelling coefficients as probability functions makes it possible to interpret our certainty about

---

5For a general and in depth overview of the Bayesian method in statistics, see e.g. McElreath (2015).

6The model description is adapted from the formulation of McElreath (2015: 304).
the estimated underlying proportions in an intuitive way, relating the coefficient’s probability to the observed data and not to a null hypothesis that assumes no underlying difference across conditions, in this case macro areas.

Coming back to the areal distribution of article inflection, Table 8.11 shows the observed counts and proportions of inflecting articles across the six macro areas in the sample.

Table 8.11: Observed proportions of inflecting articles across macro areas

<table>
<thead>
<tr>
<th>macro area</th>
<th>N inflected : N total</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>6 : 26</td>
<td>0.23</td>
</tr>
<tr>
<td>Eurasia</td>
<td>7 : 27</td>
<td>0.26</td>
</tr>
<tr>
<td>South America</td>
<td>8 : 20</td>
<td>0.40</td>
</tr>
<tr>
<td>Africa</td>
<td>12 : 29</td>
<td>0.41</td>
</tr>
<tr>
<td>Papunesia</td>
<td>13 : 31</td>
<td>0.42</td>
</tr>
<tr>
<td>Australia</td>
<td>9 : 15</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Figure 8.2 visualises the probability distributions of the proportions estimated by the model, the so-called posterior distributions. Those probability distributions, plotted as density functions, provide a more detailed insight in the degree of certainty that we have regarding the underlying proportions of inflecting articles in the six macro areas than single values with confidence intervals.

Figure 8.2: Posterior distributions for article inflection across the six macro areas
Since a probability distribution always covers an area of 1 in total, we can compare the shapes of the different curves in Figure 8.2 directly: a steeper curve corresponds to a higher degree of certainty about the underlying proportion, a flatter curve corresponds to lower probability levels for single values of the proportion and less certainty about the estimated underlying proportion. The black bar on the bottom of each distribution in Figure 8.2 indicates a 95% credible interval; it corresponds to the central 95% of the area below the curve. The credible interval shows, given prior assumptions, the observed data, and the model, the range in which the proportion coefficient falls with 95% certainty. The dot shows the median of the probability distributions of the underlying proportion; it can be taken as the best model estimate for the underlying proportion of inflecting articles. Comparing the posterior distributions in Figure 8.2, we clearly see that the ones of North America and Eurasia are very similar. Those two areas may show a little less article inflection than South America, Africa, and Papunesia, but similar proportions of inflection in all those five areas cannot be excluded with much certainty. Australia’s trend towards inflecting articles is clearly higher than the one of North America and Eurasia. In contrast, we have less certainty about an underlying difference between Australia on the one hand and Papunesia, Africa, and South America on the other. To sum up the results concerning the areal distribution of inflecting articles, we cannot exclude with absolute certainty that the proportions of inflecting areas are similar across all 6 macro areas, but given the sample, weak areal biases can be expected.

Turning to the grammatical categories that are marked on articles, Figure 8.3 shows the combinations of categories that articles inflect for in the sample. Most of the articles inflect for either number or gender, or both, and a relatively small number of articles in the sample inflects for case. This is an expected distribution, given that in the world’s languages, number and gender are more frequent nominal categories than case.

Figure 8.3: Categories of article inflection
Zooming in to categories, Figure 8.4 shows the categories for which the articles in the sample inflect across the six macro areas (only articles that inflect are included here). We can see the following areal trends: while articles inflect for number in all areas, case marking is absent in South America and Africa (and almost absent in Papunesia). Again, this is simply a consequence of case itself being less frequent in these areas. Articles that only mark number or gender are found in all areas, whereas case marking without inflection for another category is almost only found in Australia.

![Figure 8.4: Categories of article inflection in the six macro areas](image)

Another factor that may have an impact on the inflectional behavior of articles are the position of the article relative to the noun. Figure 8.5 shows the position of the article with respect to the noun and its inflectional behaviour.\(^7\) It seems that postposed (post) articles have a slightly stronger crosslinguistic preference towards inflection than preposed (pre) articles. Table 8.12 summarizes the counts and proportions of articles with inflection for those that precede and follow the noun.

---

\(^7\)The indefinite article of Sri Lanka Malay is not included here, as it appears to be flexible in preceding or following the noun.
A binomial regression model also shows that the underlying proportions of inflecting articles are very similar for both preposed and postposed articles. The fitted model is described in (19):

(19) A binomial regression that models the proportion of article inflection as a function of the position of the article. This estimates the association between the position of the article and its inflection probability.

\[(N \text{ inflected} \mid \text{trials}(N \text{ total}) - \text{position})\]

The probability distributions of the estimated proportion coefficients are shown in Figure 8.6; the credible intervals show a great degree of overlap and the areas with the highest probabilities are very close for preposed and postposed articles, so that no underlying difference can be assumed with certainty.
Figure 8.6: Posterior distributions for article inflection depending on the position

Figure 8.7 shows the proportion of inflecting and non-inflecting articles, now across the 8 different article types. Table 8.13 summarizes those counts showing the observed proportions of inflecting articles across the 8 types. Visual inspection suggests minor differences with regard to the proportion of inflecting articles across different article types; for instance, definite articles show a higher degree of inflection than indefinite articles.
Table 8.13: Observed proportions of inflecting articles across article types

<table>
<thead>
<tr>
<th>article type</th>
<th>N inflected : N total</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>indefinite</td>
<td>7 : 36</td>
<td>0.19</td>
</tr>
<tr>
<td>recognitional</td>
<td>2 : 6</td>
<td>0.33</td>
</tr>
<tr>
<td>definite</td>
<td>17 : 49</td>
<td>0.35</td>
</tr>
<tr>
<td>anaphoric</td>
<td>10 : 27</td>
<td>0.37</td>
</tr>
<tr>
<td>referential</td>
<td>3 : 6</td>
<td>0.50</td>
</tr>
<tr>
<td>inclusive-specific</td>
<td>4 : 7</td>
<td>0.57</td>
</tr>
<tr>
<td>exclusive-specific</td>
<td>8 : 12</td>
<td>0.66</td>
</tr>
<tr>
<td>nonspecific</td>
<td>4 : 5</td>
<td>0.80</td>
</tr>
</tbody>
</table>

To examine the influence of the article type on its inflectional behaviour, the following model was fitted:

\[
\text{binomial regression that models the proportion of article inflection as a function of the article type. This estimates the association between the article type and the inflection probability of the article.}
\]

\[
(N \text{ inflected } | \text{ trials(N total)} - \text{ article type})
\]

Figure 8.8 shows the probability distributions of the estimated proportions of inflecting articles across types.

Figure 8.8: Posterior distributions for article inflection depending on the article type
As can be seen in Figure 8.8, the posterior distributions for recognitional, referential, and inclusive-specific articles are extremely flat: their underlying proportions cannot be interpreted with much certainty. As for the remaining article types, we can assume that indefinite articles are less likely to inflect than anaphoric and definite articles. Exclusive-specific and nonspecific articles, on the other hand, are very likely to show more inflection than those three other types.

That definite and anaphoric articles have a stronger tendency to inflect than indefinite articles can be accounted for by their respective sources: in the definite domain, it is often demonstratives or personal pronouns which are more prone to mark especially number than the numeral ‘one’, the most frequent source for indefinite articles. As I show in section 8.1.3.2, the inflectional properties are largely restricted by the inflectional properties of their source element. It is therefore not surprising to find a higher degree of inflection with definite and anaphoric than with indefinite articles.

It is also noteworthy that there is such a big difference between the inflectional tendencies of indefinite articles on the one hand, and exclusive-specific and nonspecific articles on the other, given that the article types are semantically very similar and all belong to the indefinite domain. Especially indefinite and exclusive-specific articles often have the same diachronic source, the numeral ‘one’, which is why it is remarkable that those two article types have clearly different inflectional preferences. A potential explanation for this lies in the areal distribution of those article types. As will be shown in section 8.3.1, indefinite articles are frequent in Eurasia and North America, which were shown earlier in this section to be the two areas with the weakest trend towards article inflection. In contrast, most of the exclusive-specific articles in the sample are found in Africa, and Africa showed a significantly higher proportion of inflecting articles than both North America and Eurasia. A similar explanation applies to nonspecific articles: they are rare in the sample, and most of them occur in Papunesia. Papunesia was shown to pattern with Africa in terms of its proportion of inflected articles, so that the areal distribution may account for the differences in the inflectional behaviour similar article types in this case as well.

Thus, so far, it could be shown that probability of article inflection can partially be accounted for by the macro area. Additional differences in inflectional preferences of single article types could be related to common diachronic sources of certain article types and their geographical distribution.

Zooming in to the various categories that the 8 article types inflect for, a similar picture emerges in Figure 8.9. While there seems to be no clear trend with respect to gender and number marking, we see that case marking occurs mainly with definite and anaphoric articles from the definite domain and less so with articles from the indefinite domain. This reflects the distribution of case as a nominal category in predominantly Eurasia and Australia. Eurasia contributes to case
marking on definite articles (and to a lesser extent, indefinite ones), whereas anaphoric articles are very prominent in Australia which can explain their high proportion of case marking.

Figure 8.9: Inflectional categories of different article types

Another tendency to examine regarding inflectional preferences of articles concerns languages with more than one article. If we assume that they belong to a determiner category, we would expect different articles within single languages to show similar inflectional properties. From the sample, 31 languages have more than one article. The languages together with their inflectional properties are listed in Table A.7 in Appendix A. From those languages, both behaviors are found: articles within languages often have the same inflectional properties, but they do not need to.

Ch’ol is an example of a language with both a definite and an indefinite article which have different inflectional properties. As was shown in the preceding section, the indefinite article marks gender. The definite article, on the other hand, does not inflect. Examples (21) and (22) show this contrast.

(21) a. tyi j-k’ajty-i-b-e [jum-p’e baso]  
   ‘I ask him for a glass.’

b. tyi j-k’ajty-i-b-e [li baso]  
   PFV S:1SG-ask-TR-APPL-TR ART:DEF glass  
   ‘I ask him for the glass.’

Ch’ol (Vázquez Alvarez 2011: 247)
(22) a. ya’ añ [juñ-tiikil wiñik] there exist ART:INDEF-CL:TYIKIL man
   ‘There is a man.’

   b. ts-ach=bi majl-i [li wiñiki]
      PFV-AFF=QUOT go-itr/ipfv-o:3 ART:DEF man
      ‘It is said that the man went’

As Figure 8.10 shows, in 18 languages of the sample, all articles show the same inflectional beha-

vour. However, the articles from the other 13 languages do not behave alike with respect to inflection. This suggests a weak trend towards encoding the same categories on different articles if the language has more than one. The observed proportion of articles with the same behavior are shown in Table 8.14.

![Figure 8.10: Inflectional behavior of articles within languages](image)

Table 8.14: Observed proportion of articles within languages with the same inflectional behavior

<table>
<thead>
<tr>
<th>Inflectional behaviour</th>
<th>N same : N total</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>same</td>
<td>18 : 31</td>
<td>0.58</td>
</tr>
</tbody>
</table>

The following model was fitted to examine this trend in more detail:

(23) A binomial regression that models the proportion of articles with the same inflectional behavior with languages. This estimates the underlying probability of different articles
within languages to show the same inflectional behavior.

\[
(N \text{ same } | N \text{ total } - 1)
\]

Figure 8.11 shows the estimated probability distribution of the proportion of articles with the same inflectional behavior within languages. It shows that there may be a weak, but clearly no strong, underlying crosslinguistic trend for articles within languages to inflect for the same categories.

![Posterior distribution for articles within languages with the same inflectional behavior](image)

Figure 8.11: Posterior distribution for articles within languages with the same inflectional behavior

To conclude this section, we saw mainly areal effects that influenced the probability of an article having inflection or not, and those areal differences could also account for certain preferences across article types regarding their inflectional behavior. However, those tendencies can only account for some of the patterns attested. Therefore, I discuss two additional factors and their impact on the inflection of articles in the next section.

### 8.1.3 Factors that condition the article inflection

#### 8.1.3.1 Inflection of the noun

One could argue that marking number, gender, and case more than once in the nominal phrase would be less economic than marking it only once. Accordingly, one may expect the absence of article inflection in languages in which the noun already inflects for number and gender. On the other hand, one could also argue that marking those categories on several elements, possibly with the same exponent, reduces processing load, since the markers would signal that different words are all part of the noun phrase. In that case, one would probably expect to also find inflecting articles in languages that have nouns inflecting for those categories. Without subscribing to either views here, it is certainly plausible that the inflectional behaviour of the noun has an impact on the one of the article in addition to the factors discussed in the previous section. Examining the relation between the inflection of the noun and the inflection of the article is the aim of this section. As was already mentioned, articles are regarded as marking number, gender, and case if they do so systematically. The same applies to nouns, only that nouns are excluded for marking
gender, since gender is an inherent property of the noun and can not be marked on itself through agreement by definition. For the purposes of this section, I further restrict the marking of number and case on the noun to the contexts in which it occurs together with the article. In addition, in order to compare how often the three categories are marked, it is important to include the number of languages in the sample that feature number, gender, and case as nominal categories. Being a grammatical category of the noun does however not entail that it has to be marked on the noun or on the article; especially number and gender are marked by verbal agreement or agreement with adjectives and are neither expressed morphologically on the noun or on the article.

With this in place, Figure 8.12 presents the distributions of number, case, and gender marking on the article, noun, and on both elements as well as the number of languages in the sample that have those categories. For each of the three categories in Figure 8.12, the first bar (in red) shows how many languages of the sample have number, case, and gender as nominal categories. In relation to this figure, we see that number and case marking show very similar distributions: they are predominantly marked on the noun only (in blue), while their marking on the article (in orange) and on both the article and the noun (in green) appears less favored crosslinguistically. Gender, on the other hand, is a nominal category in roughly the same number of languages (62) in the sample as case (53). Comparing its marking on the article in Figure 8.12, we see that gender is clearly marked on the article in more languages than case.

For details on the inflectional properties of the articles in the sample, see Tables A.1 to A.6 in Appendix A.
being marked on both the article and the noun were added to both counts for article only and noun only. The data set that served as the basis for modelling the proportions of article and noun inflection is summarized in Table 8.15.

Table 8.15: Summary proportions for article and noun inflection

<table>
<thead>
<tr>
<th>category</th>
<th>N on art</th>
<th>prop on art</th>
<th>N on noun</th>
<th>prop on noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>number</td>
<td>38 : 127</td>
<td>0.30</td>
<td>7 : 127</td>
<td>0.55</td>
</tr>
<tr>
<td>case</td>
<td>13 : 53</td>
<td>0.25</td>
<td>41 : 53</td>
<td>0.77</td>
</tr>
<tr>
<td>gender</td>
<td>34 : 62</td>
<td>0.55</td>
<td>0 : 62</td>
<td>0</td>
</tr>
</tbody>
</table>

The following two models were fitted for nouns and articles to examine the association between the inflection on the noun and on the article:

(24) A binomial regression that models the proportion of inflection on the noun/article as a function of the three categories number, case, and gender. This estimates the association between the three categories and their marking on the noun/article.

1. N on noun | trials(N is category) ~ category
2. N on article | trials(N is category) ~ category

Figure 8.13 visualizes the probability distributions of different estimated proportions of noun and article inflection for number, gender, and case. We clearly see that articles are the mirror image of nouns. On the one hand, articles are significantly less likely to inflect for number and case than nouns. On the other hand, articles are significantly more likely to inflect for gender than for number and case, which can be taken as the mirror image of nouns in the sense that nouns are likely to be inflected for number and case but do not inflect for gender by definition.

Looking at the proportions of number and case inflection on nouns, Figure 8.13 confirms another trend: out of the languages that have number and/or case, we can expect to find significantly more languages that mark case on the noun than number when the article is present. This difference may have implications for the representation of case in formal syntactic theories that distinguish between a noun phrase (NP) and a determiner phrase (DP): given that case marking is crosslinguistically more frequent on the noun than on the article, the default locus for case marking should probably be the NP rather than the DP.
The fact that nouns can generally occur without an article but not vice versa suggests that the correlation between number, case, and gender marking on the noun and on the article is evidence for nominal inflection having an impact on the inflection of the article. This in turn suggests an explanation along the lines of efficiency: articles tend to mark gender over number and case as the outcome of a crosslinguistic tendency to avoid the marking on a second element in the noun phrase.

What is surprising with regard to this effect is the strength of the trend, since those three categories are often not marked separately on either articles or nouns. Instead, both often combine the expression of more than one category. As was shown in Figure 8.4, gender marking often occurs together with number marking, and if case marking occurs on the article, it is almost exclusively marked in combination with gender and/or number (except for Australia). That the trend discussed in this section is so robust can be viewed as an additional indication of a category-independent trend against repeated marking in the noun phrase.
8.1.3.2 Inflection of the article’s source element

Another main factor to influence the inflectional behaviour of articles is the inflectional behaviour of their source element, i.e. the marker that the article has developed from. In most cases, this is a demonstrative or pronoun (possessive or personal) for articles from the definite domain (cf. De Mulder & Carlier 2011), and the numeral ‘one’ for articles from the indefinite domain (cf. Heine 1997).

Three scenarios are possible with respect to the inflection of articles and their source elements: the article can either retain the inflectional morphology of its source element, simplify it which results in the loss of inflectional morphology or in the loss of form oppositions; or the article can develop new inflectional morphology that is not present on its source element. In this section, I present evidence for the first two scenarios (retaining and losing inflectional morphology) and show that there is almost no evidence for the emergence of inflectional morphology on the article.

Since diachronic evidence is not available for most of the languages in the sample, I base this section on the inflectional patterns of the article and its source element in the current stage of the languages. This is turn is based on the assumption that inflectional morphology is mostly retained on the source elements. This assumption has two consequences for the results of a synchronic comparison of the inflection on the article and on its source element: on the one hand, the source element could lose inflection that is retained on the article which I cannot distinguish from innovative inflection on the article. Since there are only two apparent exceptions which I discuss below, I do not regard potential inflection loss on the source element without its loss on the article as problematic. On the other hand, I cannot distinguish between the loss of inflection on the article and the innovation of inflection on its source element. However, the scenario of inflectional innovations on the source elements (demonstrative, adnominal possessive and numeral ‘one’) without innovations on the article is not very likely for the following reason: demonstratives, possessives, or the numeral ‘one’ occur less frequently than their article counterparts; it is not likely that the grammaticalization leading to new inflectional morphology would only affect these three elements without affecting the more frequent article.

Tables 8.16 to 8.18 show the inflectional behaviour of articles and different source elements for those articles in the sample that can be related to their source elements. Table 8.16 lists articles that originate from personal and possessive pronouns. The third column shows the categories that the article inflects for, and in the last column we see the inflection of the pronoun in the current language. Accordingly, Table 8.17 shows this for definite articles that originate from demonstratives, and Table 8.18 for articles from the indefinite domain that developed from the
numeral 'one'. In all three cases, we mainly see evidence for two of the three scenarios mentioned above: articles either keep the inflectional properties of their source elements or they simplify the system. Crucially, there are only two examples of what appears to be a morphological innovation in my sample: the indefinite articles in Cape Verde Creole and Basque (cf. Table 8.18). As I will show later in this section, those two cases are most likely no morphological innovations in a strict sense.

Table 8.16: Inflection of articles and their source elements (personal/possessive pronouns)

<table>
<thead>
<tr>
<th>language</th>
<th>article</th>
<th>ART inflection</th>
<th>pronoun</th>
<th>PRO inflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akan (ANA)</td>
<td>no</td>
<td>/</td>
<td>no (O:3SG.ANIM)</td>
<td>/</td>
</tr>
<tr>
<td>Buwal (DEF)</td>
<td>anta</td>
<td>/</td>
<td>anta (POSS:3SG)</td>
<td>/</td>
</tr>
<tr>
<td>Oko (DEF)</td>
<td>àye (sg),</td>
<td>num</td>
<td>àye (3SG),</td>
<td>num</td>
</tr>
<tr>
<td></td>
<td>âbe (pl)</td>
<td></td>
<td>âbe (3PL)</td>
<td></td>
</tr>
<tr>
<td>Pap. Malay (DEF)</td>
<td>de (3SG)</td>
<td>num</td>
<td>de (3SG)</td>
<td>num</td>
</tr>
<tr>
<td>Biak (EXSPEC)</td>
<td>-ya/-i (3SG)</td>
<td>num+gen</td>
<td>i</td>
<td>num+gen</td>
</tr>
<tr>
<td></td>
<td>(cf. Table 8.19)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warrwa (ANA)</td>
<td>kinya</td>
<td>case</td>
<td>kinya (3SG)</td>
<td>num+case</td>
</tr>
<tr>
<td>Diyari (DEF)</td>
<td>nhawu (3NF.SG)</td>
<td>num+gen+case</td>
<td>nhawu (3NF.SG)</td>
<td>num+gen+case</td>
</tr>
<tr>
<td></td>
<td>(cf. Table 8.9)</td>
<td></td>
<td>(cf. Table 8.9)</td>
<td></td>
</tr>
<tr>
<td>Cupeño (DEF)</td>
<td>pe' (sg)</td>
<td>num</td>
<td>pe' (3SG)</td>
<td>num</td>
</tr>
<tr>
<td></td>
<td>pem (pl)</td>
<td></td>
<td>pem (3PL)</td>
<td></td>
</tr>
<tr>
<td>Apinayé (DEF)</td>
<td>ja</td>
<td>/</td>
<td>ja (3SG)</td>
<td>num</td>
</tr>
</tbody>
</table>

The forms provided for articles and their source elements in Tables 8.16 to 8.18 do not always correspond to an exhaustive presentation of all exponents, but serve as an illustration for the similarity between the two elements or a subset of them.

---

9The forms provided for articles and their source elements in Tables 8.16 to 8.18 do not always correspond to an exhaustive presentation of all exponents, but serve as an illustration for the similarity between the two elements or a subset of them.
Table 8.17: Inflection of articles and their source elements (demonstratives)

<table>
<thead>
<tr>
<th>Language</th>
<th>Article</th>
<th>ART inflection</th>
<th>Demonstrative</th>
<th>DEM inflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supyire</td>
<td>-ke (C1.2.SG)</td>
<td>num+gen</td>
<td>-ṉg̱é (C1.2.SG)</td>
<td>num+gen</td>
</tr>
<tr>
<td></td>
<td>-pe (C1.5)</td>
<td>(cf. Table 8.4)</td>
<td>-mp̱é (C1.5)</td>
<td></td>
</tr>
<tr>
<td>Konso</td>
<td>-siʔ (M.SG)</td>
<td>num+gen</td>
<td>-oosiʔ (M.SG)</td>
<td>num+gen</td>
</tr>
<tr>
<td></td>
<td>-siniʔ (M.PL)</td>
<td>(m.sg)</td>
<td>-oosiniʔ (M.PL)</td>
<td></td>
</tr>
<tr>
<td>Sheko</td>
<td>-ns/hf (M.SG)</td>
<td>gen</td>
<td>y iž (M.SG)</td>
<td>num+gen</td>
</tr>
<tr>
<td></td>
<td>-in (F.SG)</td>
<td>(m.sg)</td>
<td>y īnì (F.SG)</td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td>a</td>
<td>/</td>
<td>az (NOM.SG)</td>
<td>num+gen</td>
</tr>
<tr>
<td>Armenian</td>
<td>-n</td>
<td>(m.sg)</td>
<td>-ayn (SG)</td>
<td>num+case</td>
</tr>
<tr>
<td>Bajau</td>
<td>e</td>
<td>/</td>
<td>e (DEM:DIST)</td>
<td>/</td>
</tr>
<tr>
<td>Maori</td>
<td>te (SG)</td>
<td>num</td>
<td>tēnā (SG)</td>
<td>num</td>
</tr>
<tr>
<td></td>
<td>nga (PL)</td>
<td>(M.PL)</td>
<td>ēnā (PL)</td>
<td></td>
</tr>
<tr>
<td>Alamblak</td>
<td>ind</td>
<td>num+gen</td>
<td>indar</td>
<td>num+gen</td>
</tr>
<tr>
<td>Ch’ol</td>
<td>li</td>
<td>/</td>
<td>ili</td>
<td>/</td>
</tr>
<tr>
<td>Chumash</td>
<td>lo’ka</td>
<td>/</td>
<td>lo’ (DEM:DIST)</td>
<td>/</td>
</tr>
<tr>
<td>Ute</td>
<td>‘u (S.AN.SG)</td>
<td>num+gen+case</td>
<td>‘uwā (S.AN.SG)</td>
<td>num+gen+case</td>
</tr>
<tr>
<td></td>
<td>‘umu (S.AN.PL)</td>
<td>(cf. Table 8.6)</td>
<td>‘umu (S.AN.PL)</td>
<td></td>
</tr>
<tr>
<td>Tarahumara</td>
<td>echi</td>
<td>/</td>
<td>echi</td>
<td>/</td>
</tr>
<tr>
<td>Arawak</td>
<td>li (M.SG), to (NM.SG), na (PL)</td>
<td>num+gen</td>
<td>lihi (M.SG), toho (NM.SG), naha (PL)</td>
<td>num+gen</td>
</tr>
</tbody>
</table>
Table 8.18: Inflection of articles and their source elements (numeral ‘one’)

<table>
<thead>
<tr>
<th>Language</th>
<th>Article</th>
<th>ART Inflection</th>
<th>Numeral</th>
<th>Num Inflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angolar (INDEF)</td>
<td>uⁿa</td>
<td>/</td>
<td>uma¹⁰ (sg.f)</td>
<td>num+gen</td>
</tr>
<tr>
<td>Cape Verde Creole (INDEF)</td>
<td>um (sg)</td>
<td>num</td>
<td>um</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>uns (pl)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logba (INDEF)</td>
<td>ɔkpie (cl7.sg)</td>
<td>gen</td>
<td>ikpe</td>
<td>gen</td>
</tr>
<tr>
<td>Konso (INDEF)</td>
<td>takkain (f)</td>
<td>gen</td>
<td>takka</td>
<td>gen</td>
</tr>
<tr>
<td>Hungarian (INDEF)</td>
<td>egy</td>
<td>/</td>
<td>egy</td>
<td>/</td>
</tr>
<tr>
<td>Albanian (INDEF)</td>
<td>një</td>
<td>/</td>
<td>një</td>
<td>/</td>
</tr>
<tr>
<td>Armenian (INDEF)</td>
<td>mi</td>
<td>/</td>
<td>mek</td>
<td>/</td>
</tr>
<tr>
<td>German (INDEF)</td>
<td>ein</td>
<td>gen+case</td>
<td>eins</td>
<td>gen+case</td>
</tr>
<tr>
<td>Basque (INDEF)</td>
<td>bat (nom.sg)</td>
<td>num+case</td>
<td>bat</td>
<td>case</td>
</tr>
<tr>
<td></td>
<td>batzuk (nom.pl)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barwar (INDEF)</td>
<td>xa (m)</td>
<td>gen</td>
<td>xa (m)</td>
<td>gen</td>
</tr>
<tr>
<td></td>
<td>ḍa (f)</td>
<td></td>
<td>ḍa (f)</td>
<td></td>
</tr>
<tr>
<td>Kharia (INDEF)</td>
<td>moɲ</td>
<td>/</td>
<td>muɁu (hum)</td>
<td>gen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>moɲ (nhum)</td>
<td></td>
</tr>
<tr>
<td>Kurtöp (INDEF)</td>
<td>the</td>
<td>/</td>
<td>the</td>
<td>/</td>
</tr>
<tr>
<td>Tamil (INDEF)</td>
<td>oru</td>
<td>/</td>
<td>onru</td>
<td>/</td>
</tr>
<tr>
<td>SL Malay (INDEF)</td>
<td>hatthu</td>
<td>/</td>
<td>satthu</td>
<td>/</td>
</tr>
<tr>
<td>Agul (INDEF)</td>
<td>sa</td>
<td>/</td>
<td>sa</td>
<td>/</td>
</tr>
<tr>
<td>Palula (exspec)</td>
<td>āak/āa</td>
<td>/</td>
<td>āak/āa</td>
<td>/</td>
</tr>
<tr>
<td>language</td>
<td>article</td>
<td>ART inflection</td>
<td>numeral</td>
<td>NUM inflection</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>----------------</td>
<td>---------</td>
<td>----------------</td>
</tr>
<tr>
<td>Komnzo</td>
<td>nā</td>
<td>/</td>
<td>nābi</td>
<td>/</td>
</tr>
<tr>
<td>Sye</td>
<td>hai</td>
<td>/</td>
<td>hai(ten)</td>
<td>/</td>
</tr>
<tr>
<td>Teiwa</td>
<td>nuk</td>
<td>/</td>
<td>nuk</td>
<td>/</td>
</tr>
<tr>
<td>Chocholtec</td>
<td>naa</td>
<td>/</td>
<td>naa</td>
<td>/</td>
</tr>
<tr>
<td>Tepehua</td>
<td>tam</td>
<td>/</td>
<td>laqa-tam (CL:GEN-ONE)</td>
<td>gen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>puma-tam (CL:HUM-ONE)</td>
<td></td>
</tr>
<tr>
<td>Mapudungun</td>
<td>kiñe</td>
<td>/</td>
<td>kiñe</td>
<td>/</td>
</tr>
<tr>
<td>Nhengatu</td>
<td>yepe</td>
<td>/</td>
<td>yepe</td>
<td>/</td>
</tr>
<tr>
<td>Parecis</td>
<td>hatya</td>
<td>/</td>
<td>hatita</td>
<td></td>
</tr>
<tr>
<td>Wayuu</td>
<td>wanee</td>
<td>/</td>
<td>wanee</td>
<td>/</td>
</tr>
</tbody>
</table>

The two potential exceptions to this trend both involve an indefinite article that can be marked for number. Those exceptions are rather only apparent because the source element, the numeral ‘one’, is simply not attested with this plural marker due to the obvious semantic incompatibility between the numeral and the plural. On the basis of positive evidence only, it is hard to distinguish between a semantic restriction that leads to the form not being attested, and a morphological restriction in the strict sense. In addition, the indefinite article un in Cape Verde Creole relates to the numeral ‘one’ in the language, but is also lexically based on the indefinite article in Portuguese, which is can occur with plural nouns and which is morphologically marked for plural as well. So the plural marking in Cape Verde Creole, shown in (25b) cannot really be viewed as innovative.

(25) a. N ten [un duensa] ki’ N dja ara ku el ....
   I have ART:INDEF.SG disease COMP I PFV bother with it ...
   ‘I have a disease that bothers me ...’ Cape Verde Creole (Baptista 2002: 31-32)

b. Toki’ N ta ten [uns problema] ...
   when I NPST have ART:INDEF.PL problem ...
   ‘When I have some problems ...’ Cape Verde Creole (Baptista 2002: 27)

10 What is given here as the (lexical) source element is not the numeral ‘one’ in Angolar (Creole), but the Portuguese indefinite article.
In Basque, the plural marking on the indefinite article, as shown in (26b) is perhaps not innovative in the strict sense either. The segment -k generally occurs with nouns and adjectives to mark plural in the absolutive and ergative (Hualde & Ortiz de Urbina 2003: 173). Also the inclusive specific article a in Basque is followed by -k in these contexts. In addition, the plural marker only occurs on the right edge of the noun phrase (Hualde & Ortiz de Urbina 2003: 113). Thus, the fact that the indefinite article occurs with the plural marker is not due to innovative morphology, but a consequence of an independent property of the plural marker in the language.

(26) a. 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)

    b. zergatik ez erregutu hari [bitxi bat-zuk] utz ziezakion?
    why not beg him jewel ART:PRES-PL lend AUX.S.COMP
    ‘Why not beg her to lend her some jewels?’ Basque (Hualde & Ortiz de Urbina 2003: 503)

To conclude, the data in Tables 8.16 to 8.18 showed that the inflectional potential of an article is generally constrained by the inflectional morphology of its source element, independently from the type of source element. We almost exclusively find changes that correspond to the loss of formal oppositions leading to the loss of inflection on the article. This strongly suggests that the categories that articles inflect for are restricted by the properties of their source elements to begin with, since simplification but (almost) no innovation could be observed.

8.2 Article paradigms

This section is concerned with article paradigms. I discuss two types of paradigmatic relations concerning articles: the first part deals with the form relation of different articles within languages, and the second part presents the paradigm of the definite article in Mokpe (Bantu, Cameroon). The latter is an example of a complex article paradigm, showing how different exponents of a single article can interact.

8.2.1 One article or many?

In the world’s languages, we find article systems with more than one article, with one of the exponents being the morphological basis for the exponent of the other article. This is only attested with certain combinations of article types: exclusive-specific and definite articles, as well exclusive-specific and nonspecific articles.
Biak has an exclusive-specific article and a definite article (and also a nonspecific article which is not relevant here). The specific article has the exponents as shown in Table 8.19 and attaches to the last element of the noun phrase which is the noun by default (van den Heuvel 2006: 204).

Table 8.19: Biak specific article exponents

<table>
<thead>
<tr>
<th>article exponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>singular</td>
</tr>
<tr>
<td>dual</td>
</tr>
<tr>
<td>paucal</td>
</tr>
<tr>
<td>plural</td>
</tr>
<tr>
<td>-ya/-i</td>
</tr>
<tr>
<td>-su-ya/i</td>
</tr>
<tr>
<td>-sko-ya/i</td>
</tr>
<tr>
<td>-s-ya/i</td>
</tr>
</tbody>
</table>

The exponents of the exclusive-specific article are also used to form the definite article. The latter formally consists of the marker *an* and the exclusive-specific article that attaches to *an*. In example (27), we see the exclusive-specific article *ya* attached to a noun and an adjective. In (28), *ya* forms the definite article *anya* together with the marker *an*.

(27) ya-fáru man-koko=ya kukru ikák ve=ba=ya 1sg-tell bird-chicken=ART:SPEC with snake REL-big-ART:SPEC 'I tell about a bird and a big snake.' Biak (van den Heuvel 2006: 60)

(28) indya fyanu [rofan anya] i-fnovku mankroder=i so feed.3sg dog ART:DEF 3sg-with frog=ART:SPEC 'So he took care of the dog, together with a frog.' Biak (van den Heuvel 2006: 205)

The other cases of article exponents forming the base of other articles are the exclusive-specific articles in Q’anjobal (cf. section 5.2.1), Lakota (cf. section 5.2.3), and Crow (cf. 2.2.4), whose forms are repeated in Table 8.20 together with the forms of the nonspecific articles. For Q’anjobal, it was shown in section 5.2.4 that the nonspecific article originates from an irrealis marker that combined with the former indefinite article *jun* and conventionlized into a nonspecific article. Whether a similar scenario is also responsible for the formal relation of the exclusive-specific and the nonspecific articles in Crow and Lakota is unclear at this point; it seems plausible, though, that the nonspecific articles developed from the combination of an indefinite article and some other marker that restricted the combination of markers to nonspecific contexts.

---

11van den Heuvel (2006: 68) notes that "-ya can be used in all positions, whereas the use of -i is restricted to prepausal position".
Table 8.20: Specific and nonspecific article exponents

<table>
<thead>
<tr>
<th></th>
<th>exclusive-specific article</th>
<th>nonspecific article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q’anjobal</td>
<td>jun</td>
<td>jun-oq</td>
</tr>
<tr>
<td>Lakota</td>
<td>wā</td>
<td>wā-ži</td>
</tr>
<tr>
<td>Crow</td>
<td>-m</td>
<td>-eem</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>inclusive-specific article</th>
<th>nonspecific article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongan</td>
<td>(h)e</td>
<td>ha</td>
</tr>
<tr>
<td>Blackfoot</td>
<td>animate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-(w)a (sg), -(y)i (pl)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>inanimate</td>
<td>-i</td>
</tr>
<tr>
<td></td>
<td>-iksi (sg), -itsi (pl)</td>
<td></td>
</tr>
<tr>
<td>Siar Lak</td>
<td>animate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ep</td>
<td>ti</td>
</tr>
<tr>
<td></td>
<td>inanimate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>ta</td>
</tr>
</tbody>
</table>

Table 8.20 also shows the specific and nonspecific articles in Tongan (cf. section 5.2.2), Blackfoot (Frantz 1991: 9-11), and Siar Lak (Rowe 2005: 16). These three article pairs do not show any obvious form relation. This shows that the exponent of the nonspecific article does not necessarily depend on the exponent of the specific article.

Nevertheless, the formal relation between a number of certain article pairs is not random: the exclusive-specific and nonspecific article (Q’anjobal, Lakota, Crow) as well as the exclusive-specific and definite article (Biak). That the exponent of the exclusive-specific article forms the basis for the definite article in Biak is plausible in a scenario in which the ambiguity between a former inclusive-specific and definite referents was resolved by an additional marker in definite contexts, *an*, that conventionalized and led to the development of a definite article, consisting of *an* and the exponent of the later exclusive-specific article.

As was shown for nonspecific articles in section 5.2.4, their development often involves the combination of a former indefinite article with an irrealis marker (or perhaps another type of marker) that restricts this indefinite article to nonspecific contexts. Even though the nature of this additional marker may differ, the fact that we find a number of exclusive-specific and nonspecific articles for which the exponent of the former corresponds to the basis of the exponent of the latter is evidence that the diachronic development for nonspecific articles proposed in section 5.2.4 is plausible. Moreover, it accounts for the different properties of the article pairs in Tongan, Blackfoot, and Siar Lak, all of which have an article system consisting of an inclusive-specific article and a nonspecific article. These systems probably did not have an indefinite article.
from which the nonspecific article could have developed, which can explain why the exponents of these two articles are not formally related.

Another issue related to article paradigms is whether or not certain similar markers should be treated as different exponents of a single article or as different articles. What I treat as single article is based on the referential functions of the markers. Thus, forms that share their referential function and additionally distinguish between values of other categories are analyzed as exponents of the same article. These additional values can be nominal categories such as number, gender (and animacy), and case marked on the article, which seem to be uncontroversially treated as inflection of a single article anyway. In addition, across the languages of the world we find that articles can be marked for other categories, e.g. deixis. Especially if categories that interact with referentiality or that are closely related are marked additionally, article exponents tend to be treated as different articles rather than forms of a single article marked for e.g. deixis or visibility, as was shown for Halkomelem (cf. section 6.2.2), Baure (cf. section 6.2.3) and Macedonian (cf. section 8.1.1). The paradigm of the referential article in Halkomelem is repeated in Table 8.21 for convenience.

Table 8.21: Halkomelem article paradigm

<table>
<thead>
<tr>
<th></th>
<th>nonfeminine</th>
<th>feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROX &amp; VIS</td>
<td>tə</td>
<td>thə</td>
</tr>
<tr>
<td>PROX &amp; INV</td>
<td>kʷθə</td>
<td>kʷlə</td>
</tr>
<tr>
<td>REM</td>
<td>kʷɛ</td>
<td>kʷsə</td>
</tr>
</tbody>
</table>

Because the additional values of proximity, visibility, and remoteness do not directly translate into referential values and are not necessarily interpreted as such in all contexts, I treat systems with markers similar to the ones in Halkomelem as exponents of a single article.

### 8.2.2 A complex article paradigm in Mokpe

In section 4.1.2, we saw that the definite article in Mokpe (Bantu, Cameroon) has a rather complex paradigm. In this section, I present the paradigm in more detail and propose a scenario for its development. Table 4.1, repeated as Table 8.22, illustrates the article exponents for nouns of the different classes. With regard to the article, this paradigm can be viewed as merging two layers of article exponents: an older layer that corresponds to the tonal exponent, and a more recent layer which is the segmental marker.
The nouns of classes 1 to 10 are singular-plural pairs, while the nouns in classes 14 and 19 are singular forms with corresponding plural forms mostly in class 10. Each noun class is represented by a noun with a high and low tone in the first tone-bearing unit, which is affected by the tonal article exponent. What is marked in Table 8.22 are those classes which do not combine with the high tone exponent. Class 1 and 9 occur with è, while nouns of class 10 combine with the marker í. I show later in this section that í is not an exponent of the definite article and can occur in both definite as well as in indefinite contexts.

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 that already have an initial high tone. This effect is represented as vowel-lengthening here.

---

<table>
<thead>
<tr>
<th>class</th>
<th>noun</th>
<th>article+noun</th>
<th>class</th>
<th>noun</th>
<th>article+noun</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>singular</td>
<td></td>
<td></td>
<td>plural</td>
</tr>
<tr>
<td>1</td>
<td>mòtò</td>
<td>è mòtò</td>
<td>2</td>
<td>wátò</td>
<td>wátò</td>
</tr>
<tr>
<td></td>
<td>ñmánà</td>
<td>è ñmánà</td>
<td></td>
<td>wánà</td>
<td>wánà</td>
</tr>
<tr>
<td>3</td>
<td>móléli</td>
<td>móléli</td>
<td>4</td>
<td>mélélí</td>
<td>mélélí</td>
</tr>
<tr>
<td></td>
<td>mólëngò</td>
<td>mólëngò</td>
<td></td>
<td>mélëngò</td>
<td>mélëngò</td>
</tr>
<tr>
<td>5</td>
<td>likálà</td>
<td>likálà</td>
<td>6</td>
<td>màkálà</td>
<td>màkálà</td>
</tr>
<tr>
<td></td>
<td>likpà</td>
<td>likpà</td>
<td></td>
<td>màkpà</td>
<td>mà:kpà</td>
</tr>
<tr>
<td>7</td>
<td>ètáŋgùlē</td>
<td>ètáŋgùlē</td>
<td>8</td>
<td>bètáŋgùlē</td>
<td>bètáŋgùlē</td>
</tr>
<tr>
<td></td>
<td>èlélé</td>
<td>èlélé</td>
<td></td>
<td>bèlélè</td>
<td>bèlélè</td>
</tr>
<tr>
<td>9</td>
<td>mbèžà</td>
<td>è mbèžà</td>
<td>10</td>
<td>í mbèžà</td>
<td>í mbèžà</td>
</tr>
<tr>
<td></td>
<td>mbólì</td>
<td>è mbólì</td>
<td></td>
<td>í mbólì</td>
<td>í mbólì</td>
</tr>
<tr>
<td>14</td>
<td>wôzò</td>
<td>wôzò</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>wôngò</td>
<td>wôngò</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>ènòní</td>
<td>ènòní</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dzìŋgù</td>
<td>dzìŋgù</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

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

13 In 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.
As mentioned in the beginning of this section, the tonal article exponent most probably represents an older article system. It is more pervasive and the only marking for definite referents mentioned in Atindogbe (2013).

Even though there is no reason that would speak against referential markers being expressed by tones, tonal article exponents are crosslinguistically rare. To the best of my knowledge, the other only reported case is found in Ewondo (Lyons 1999: 65f), which is 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. In Mokpe, this process is pervasive with prepositions and other function words that often only consist of a single vowel and that can be reduced to a tone which associates to the first tone-bearing unit of the following word.

Possessives and demonstratives show an interesting effect with respect to the two article exponents. The co-occurrence restrictions for the tonal and segmental exponents with possessives and demonstratives are different in Mokpe. Examples (29) and (30) illustrate that the segmental article è cannot co-occur with either demonstratives or possessive markers in the noun phrase.

(29) (*è) ónó (*è) mòtò
\textit{ART:DEF DEM.CL1 ART:DEF child.CL1}
\textit{this child} Mokpe (prim. data)

(30) (*è) mòtò wà-mi
\textit{ART:DEF child.CL1 CL1-POS:1SG}
\textit{my child} Mokpe (prim. data)

The tonal article exponent, on the other hand, can be shown to occur with both demonstratives and possessives. Demonstratives precede the noun, while possessives follow it. Assuming that the anchor of the tonal article is the noun phrase and not the noun, one would expect to find it on the first tone-bearing unit of the first element in the noun phrase. Demonstratives preceding the noun, we would expect to find that all demonstrative markers of the noun classes that take the tonal article have an initial high tone, and that the high tone no longer surfaces on the noun. Possessive pronouns, on the other hand, follow the noun. Accordingly, the high tone is expected to be realized on the noun. Table 8.23 shows that this is indeed what we find for the tonal article exponent. On the left, we see the proximal demonstrative in combination with a noun, and on the right, possessives are illustrated by the possessive form of the first person singular. We see that all demonstratives of the noun classes taking the tonal article have an initial high tone, while

\footnote{The phonological assimilation of preposed articles is common across the world’s language, a well-known example being the definite article \textit{le/la} in French occurring as shortened \textit{l’} before vowel-initial nouns.}
demonstratives of class 1 and 9 have an initial low tone. Also, the noun is no longer affected by the tone if the demonstrative is present. In combination with possessives, on the other hand, the tone is marked on the noun, as it is the first element in the noun phrase. As for the segmental article è, we see in Table 8.23 that it cannot be applied in the presence of either a demonstrative or a possessive.

Table 8.23: The definite article in Mokpe with demonstratives and possessives

<table>
<thead>
<tr>
<th>class</th>
<th>this</th>
<th>noun</th>
<th>my</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(*è) ónó</td>
<td>mòtò ‘this man’</td>
<td>(*è) mòtò</td>
</tr>
<tr>
<td>2</td>
<td>wànù</td>
<td>wàtò ‘these women’</td>
<td>wàtò</td>
</tr>
<tr>
<td>3</td>
<td>mónò</td>
<td>mòléléí ‘this food’</td>
<td>mòléléí</td>
</tr>
<tr>
<td>4</td>
<td>mènè</td>
<td>mèkò ‘these plantains’</td>
<td>mèkò</td>
</tr>
<tr>
<td>5</td>
<td>lìnì</td>
<td>likálà ‘this bridge’</td>
<td>likálà</td>
</tr>
<tr>
<td>6</td>
<td>mànù</td>
<td>màkálà ‘these bridges’</td>
<td>màkálà</td>
</tr>
<tr>
<td>7</td>
<td>ènè</td>
<td>èlèlà ‘this duck’</td>
<td>èlèlà</td>
</tr>
<tr>
<td>8</td>
<td>wènè</td>
<td>wèlélà ‘these ducks’</td>
<td>wèlélà</td>
</tr>
<tr>
<td>9</td>
<td>(*è) ènè</td>
<td>mbèzhà ‘this boy’</td>
<td>(*è) mbèzhà</td>
</tr>
<tr>
<td>10</td>
<td>(*ì) ínì</td>
<td>mbèzhà ‘these boys’</td>
<td>(*ì) mbèzhà</td>
</tr>
<tr>
<td>14</td>
<td>wònò</td>
<td>wòʒó ‘this face’</td>
<td>wòʒó</td>
</tr>
<tr>
<td>19</td>
<td>ínì</td>
<td>inòñì ‘this bird’</td>
<td>inòñì</td>
</tr>
</tbody>
</table>

This shows that even though the two markers are exponents of a single definite article based on their distribution in definite contexts (cf. section 4.1.2), there seems to be no requirement for the two exponents to behave alike with respect to their co-occurrence with demonstratives and possessives.

Concerning the development of this article paradigm, there is no evidence for the source of the tonal article. However, it can be assumed to be the older article exponent that, for some reason, was no longer available for nouns of class 1, 9, and 10 at a certain point in time. This in turn could have led to the development of the segmental article exponent è. The latter can be traced back to the agreement marker of class 9, occurring in the preverbal domain and cross-referencing the subject. An example is given in (31):

(31) è mbúndá è timbá ndóló  
\textbf{ART:DEF} friendship.cl9 cl9 become love  
‘The friendship became love.’ Mokpe (prim. data)

There are two arguments to explain why the agreement marker of class 9 developed into an article exponent instead of the agreement marker á of class 1. Firstly, class 9 seems to be the most
productive noun class in Mokpe; loan words as well as ad-hoc uses of English and French nouns consistently fall into this class. As a result, class 9 nouns probably also have a high token frequency in use, which would already make them more accessible for the development into an article than less frequent forms. Secondly, it is plausible that the referents of class 1 nouns, human referents in most cases, are interpreted as highly discourse-prominent and as definite by default, making additional marking as definite less necessary than the marking of other types of referents. Once the marker è was established as an adnominal definiteness marker with class 9 nouns, the marker may have been less associated with its earlier function of marking the gender class 9, giving way to be extended to class 1 nouns. That this is a plausible scenario is reflected in the current paradigm: the agreement marker of class 1 is not attested as an article exponent.

Nouns of class 10, as shown in table 8.22, neither combine with the tonal nor the segmental article exponent but are marked as plural nouns with í. Since class 9 and 10 nouns do not have a noun class prefix that formally distinguishes number, the plural agreement marker of class 10, otherwise used with verbs, occurs adnominally to mark the nouns as plural forms. Its use as agreement marker for nouns of class 10 is illustrated in (32).

(32) [í ñòkù] í lijà ó wàngá
   cl10 elephant.cl9/10 cl10 live in forest
   'Elephants live in the forest.' Mokpe (prim. data)

In contrast to its singular counterpart è which occurs adnominally as a definite article, the marker í does not encode definiteness but marks plural. The referent type of the noun is determined by the context alone. Example (33) shows the noun kémà of class 10 marked by í. The referent of the noun is interpreted as specific from the context because it is not identifiable by the hearer. In utterances like (34), the referents of í ngòndò ‘girls’ can only be interpreted as definite according to my informants. That the plural class 10 marker is also compatible with nonspecific contexts is shown in (35).

(33) è bóndà na ðèlì, n-ɛ́ní [í kémà] ó ndàwò
   the time I came 1sg-see cl10 monkey.cl10 in house.cl2
   ‘When I came (home), I saw (some) monkeys in the house.’ Mokpe (prim. data)

(34) [í ngòndò] dzàngóg dzó ámbèlè ówòkà
   cl10 girl.cl10 your 2sg wait outside
   ‘The girls are waiting outside.’ Mokpe (prim. data)

(35) wúnjá wòkó nà kànë lôβà [í mbúndà]
   day one I pray have cl.10 friend.cl10
   ‘I wish to have friends one day.’ Mokpe (prim. data)
Nevertheless, the plural marker *í* shows some effects related to definiteness. As is shown in examples (36) and (37), existential constructions that force a non-definite interpretation of the referent do not allow for the use of *í*. This makes the noun ambiguous between a singular and plural interpretation:

(36) ó wéli [(‘í) kémà] ó wàŋgá there be cl10 monkey.cl10 in forest ‘There are / is a monkey(s) in the forest.’ Mokpe (prim. data)

(37) ó wéli [(‘í) ŋgòndɔ̀] jò ámbɛ̀lɛ̂ ówókà there be cl10 girl.cl10 2sg wait outside ‘There are girls / a girl waiting for you outside.’ Mokpe (prim. data)

The Mokpe article paradigm showed two general properties of articles and article paradigms. Firstly, the two article exponents in Mokpe stress the independence between articles as referential markers and the forms that these markers have. In addition, Mokpe uses the marker *í* to indicate plural (class 10). Even though it is similar to the segmental article exponent é, it does not mark definiteness and cannot be considered as an article exponent for nouns of class 10. Secondly, we saw that the two article exponents showed a distinct behaviour with demonstratives and possessives. This provides strong evidence against a functional motivation of the co-occurrence restriction of the definite article and possessives and demonstratives. While it may be less surprising that we find variation with respect to this parameter across languages, Mokpe shows that different exponents of the same definite article in a single language can also differ in this respect.

### 8.3 Article types

#### 8.3.1 Areal trends

In the sample of 148 articles from 113 languages from the world, the frequency distributions of the 8 article types are presented in Figure 8.14: definite (def), anaphoric (ana), recognitional (recog), exclusive-specific (exspec), nonspecific (nspc), indefinite (indef), inclusive-specific (inspec), and referential (ref) articles. We see that definite articles correspond to the most frequent article type, confirming the trend observed in Dryer (1989, 2014). However, indefinite articles do not appear to be much less frequent than definite articles. Especially given such a more fine-grained distinction of article types, definite, indefinite and anaphoric articles are all relatively frequent compared to the remaining types. Table 8.24 summarizes the counts together with the proportions for each article type in the sample.
Again, a binomial regression model can shed more light onto the underlying proportions of different article and making their differences more comparable. The model fitted is described in (38):

(38)  A binomial regression that models the proportions of the 8 article types. This estimates the underlying probabilities of the different article types.

\[(N \text{ type} \mid \text{trials}(N \text{ total}) - \text{article type})\]

Figure 8.15 shows the estimated underlying probability distributions for each article type on the basis of their distribution in the sample. We can say with certainty that definite, indefinite, and anaphoric articles are more frequent crosslinguistically than the other article types. From those
minor article types, exclusive-specific articles can be assumed to occur more frequently across languages than nonspecific and referential articles, but a very clear trend is not apparent. It is also very certain that definite articles are indeed more frequent than anaphoric articles. Clear predictions concerning the underlying frequencies of indefinite and definite articles, however, cannot be made. The model does predict a slight crosslinguistic preference for languages to have definite articles over indefinite articles, but the confidence intervals as well as the density probability functions show a great deal of overlap, which is why we cannot exclude that both article types are equally frequent in the world’s languages.

Figure 8.15: Posterior distributions for the frequencies of article types

Figure 8.16 shows the distribution of article types across the six macro areas. We can observe three major areal trends: Africa and North America show a relatively high proportion of definite articles (red). Eurasia and South America, on the other hand, feature a large number of indefinite articles (green), and Australia stands out for its high proportion of anaphoric articles (orange).
Turning to the proportion of definite articles across the 6 macro areas, Table 8.25 below summarizes their counts in the sample together with their proportions.

<table>
<thead>
<tr>
<th>macro area</th>
<th>N def : N total</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>16 : 29</td>
<td>0.55</td>
</tr>
<tr>
<td>North America</td>
<td>13 : 26</td>
<td>0.5</td>
</tr>
<tr>
<td>Eurasia</td>
<td>7 : 27</td>
<td>0.26</td>
</tr>
<tr>
<td>Papunesia</td>
<td>8 : 31</td>
<td>0.26</td>
</tr>
<tr>
<td>South America</td>
<td>4 : 20</td>
<td>0.2</td>
</tr>
<tr>
<td>Australia</td>
<td>1 : 15</td>
<td>0.07</td>
</tr>
</tbody>
</table>

To compare the proportion of definite articles across macro areas, the following model was fitted:

\[(N \text{ definite} | \text{trials}(N \text{ total}) \sim \text{macro area})\]

Figure 8.17 shows the estimated probability distributions of the occurrence of definite articles for each macro area. Australia clearly has the lowest estimated proportion of definite articles, we can be certain that this proportion is lower for Australia than for the other areas (except South America).
Similarly to the observed proportion, the model does not predict significant differences between South America, Papunesia, and Eurasia. We can assume that definite articles really are equally frequent, or rather infrequent, in those three macro areas. The two macro areas with the highest proportions of definite articles are Africa and North America, definite articles in those two areas are estimated to be equally frequent. While an underlying difference between North America and Eurasia is less certain, we can assume that languages in Africa tend to have more definite articles than languages in the other macro areas (except North America, which is similar to Africa in this respect). An evident explanation for this preference of definite articles in Africa is not at hand, and it would go beyond the purposes of the present study to investigate this trend in more detail. It may be promising to examine the relation between the presence / absence of definite articles with other grammatical phenomena such as word order, case marking, nominal classes and cross-reference on other nominal and verbal elements, obviative marking, and focus marking.

![Figure 8.17: Posterior distributions of definite articles across macro areas](image)

The second trend that emerged from the distribution presented in Figure 8.16 was the high proportion of indefinite articles in Eurasia and South America. The distribution of indefinite articles across macro areas is repeated in Table 8.26.
Table 8.26: Observed proportions of indefinite articles across macro areas

<table>
<thead>
<tr>
<th>Macro Area</th>
<th>N Indef : N Total</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eurasia</td>
<td>13 : 27</td>
<td>0.48</td>
</tr>
<tr>
<td>South America</td>
<td>8 : 20</td>
<td>0.4</td>
</tr>
<tr>
<td>North America</td>
<td>7 : 26</td>
<td>0.27</td>
</tr>
<tr>
<td>Papunesia</td>
<td>6 : 31</td>
<td>0.19</td>
</tr>
<tr>
<td>Africa</td>
<td>2 : 29</td>
<td>0.07</td>
</tr>
<tr>
<td>Australia</td>
<td>0 : 15</td>
<td>0</td>
</tr>
</tbody>
</table>

Again, a binomial model is fitted in order to examine the areal biases for the frequency of indefinite articles.

\[(40)\] A binomial regression that models the proportion of indefinite articles as a function of macro areas. This estimates the association between the frequency of the indefinite articles and the macro areas.

\[(N \text{ indefinite} \mid \text{trials}(N \text{ total}) \sim \text{macro area})\]

Figure 8.18 shows the probability distributions of the estimated proportions of indefinite articles in the six macro areas.

Figure 8.18: Posterior distributions of indefinite articles across macro areas
As expected we see a sort of a mirror image to the areal biases for definite articles shown in Figure 8.17: It is very likely that we find less languages with indefinite articles in Africa (and also in Australia) than in Eurasia, and less likely so, in South America. As for the other areas, despite the observed tendencies, we cannot be certain about clear differences: the frequency of indefinite articles in Eurasia and South America is very similar. North America has a somewhat lower estimated underlying proportion of indefinite articles than the other two areas, but we cannot be certain to find a significant difference. The same holds for Papuasia: it is likely that Papuasia has less languages with definite articles than Eurasia, but there is no other clear difference given the distributions in the sample and the model. Considering the overall preferences, we do however again see that there is an areal bias and that indefinite articles, similarly to definite articles, are not equally likely to occur in different macro areas. Again, a preference for or against indefinite articles can probably only be accounted for by its relation to other grammatical properties.

A possible hint pointing towards the preference of indefinite articles over, e.g., definite ones, in Eurasia is definiteness or specificity based differential object marking (DOM) that seems to be relatively common in a number of Iranian and Turkic languages as opposed to animacy based DOM that seems to be common in Indo-Aryan languages. Table 8.27 shows number of languages with DOM in the database of Sinnemaki (2014). At this point, those numbers are very small and can only be taken impressionistically, but they pattern with the observation that we rather find indefinite marking in Turkic and Iranian languages, but less so in Indo-Aryan languages.

<table>
<thead>
<tr>
<th>DOM type</th>
<th>Indo-Aryan</th>
<th>Iranian</th>
<th>Turkic</th>
</tr>
</thead>
<tbody>
<tr>
<td>animacy</td>
<td>2</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>animacy-definiteness</td>
<td>5</td>
<td>/</td>
<td>1</td>
</tr>
<tr>
<td>definiteness</td>
<td>/</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>no DOM</td>
<td>/</td>
<td>1</td>
<td>/</td>
</tr>
</tbody>
</table>

Table 8.27: Differential object marking in Indo-Aryan, Iranian, and Turkic (Sinnemaki 2014)

A system with definiteness or specificity based DOM already incorporates the marking of referentiality to a certain degree, as is shown in (41) below for Turkish:

(41) a. Mary kek-i yap-ti
     Mary cake_ace make-pst
     ‘Mary baked the cake’

b. Mary bir kek yap-ti
   Mary art:indef cake make-pst
   ‘Mary baked a cake’

   Turkish (Hedberg et al. 2009: 13)
Given that referents in the subject position can be assumed to be definite in most cases, and given that definite objects are marked as opposed to indefinite objects that are left unmarked, this type of DOM may cause a bias against the development of definite articles, which would generally the most likely article to be developed. Thus, because definiteness is already marked in at least certain relevant contexts, the bias against the development of definite articles may result in a bias towards the development of indefinite articles. Since the relation between DOM, or rather differential argument marking in general, and the availability of articles would go beyond the scope of this study, future research is required to shed more light on the relation between those two grammatical phenomena.

The third trend that emerged in Figure 8.16 concerned the presence of anaphoric articles being preferred in Australia compared to the other macro areas. Table 8.28 repeats the counts of anaphoric articles across the six macro areas together with their proportions.

Table 8.28: Observed proportions of anaphoric articles across macro areas

<table>
<thead>
<tr>
<th>Macro area</th>
<th>N ana : N total</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>4 : 29</td>
<td>0.14</td>
</tr>
<tr>
<td>Eurasia</td>
<td>3 : 27</td>
<td>0.11</td>
</tr>
<tr>
<td>Papunesia</td>
<td>4 : 31</td>
<td>0.13</td>
</tr>
<tr>
<td>Australia</td>
<td>11 : 15</td>
<td>0.73</td>
</tr>
<tr>
<td>N America</td>
<td>2 : 26</td>
<td>0.08</td>
</tr>
<tr>
<td>S America</td>
<td>3 : 20</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Even though a clear difference between Australia and all other macro areas is evident, a binomial model was fitted for the sake of completeness.

(42) A binomial regression that models the proportion of anaphoric articles as a function of macro areas. This estimates the association between the frequency of the anaphoric articles and the macro areas.

\[(N \text{ anaphoric} | \text{ trials}(N \text{ total}) - \text{ macro area})\]

Figure 8.19 shows the estimated probability distributions for the proportion of anaphoric articles in the six macro areas. While there is no reliable difference across areas in general, Australia clearly stands out in showing a higher proportion of anaphoric articles than all other macro areas. Even though its credible interval overlaps with the one of South America, the distributions discussed in this section so far all indicate that Australia shows a very different trend with regard to articles than all other areas. This can either be formulated as a very strong tendency towards anaphoric articles and against articles in the indefinite domain (Australia only features articles in
the definite domain in the sample). Or, taking a more conservative approach that does not treat adnominal anaphoric markers as articles but as demonstratives, the conclusion would be that Australia, as opposed to all other areas of the world, has a very strong preference against articles in general.

![Figure 8.19: Posterior distributions of anaphoric articles across macro areas](image)

A partial explanation that covers the preference towards anaphoric articles in Australia can be related to the personal and demonstrative pronoun systems. Many of the Australian languages in the sample either lack distinct third person pronouns (and use demonstratives) or the pronouns are described as being rarely used in speech.\footnote{In the WALS chapter on third person pronouns and demonstratives (Bhat 2013), the data points for Australia also show a high number of formally related third person pronouns and demonstratives.} This is summarized in Table 8.29. Out of the 11 languages with anaphoric articles in Australia, three languages do not have a third person pronoun that is formally distinct from demonstratives to begin with, and for 4 additional languages, the grammars report that third person pronouns either have defective paradigms and/or are rarely used in speech. 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 8.29: The relation of third person pronouns and demonstratives in Australian languages

<table>
<thead>
<tr>
<th>language</th>
<th>properties of third person pronouns</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mangarrayi</td>
<td>no third person pronoun distinct from the demonstrative</td>
<td>(Merlan 1989: 52)</td>
</tr>
<tr>
<td>Wardaman</td>
<td></td>
<td>(Merlan 1994: 107f)</td>
</tr>
<tr>
<td>Warrwa</td>
<td></td>
<td>(McGregor 1994: 20)</td>
</tr>
<tr>
<td>Jingulu</td>
<td>pronoun forms only available in ACC and GEN</td>
<td>(Pensalfini 1997: 242)</td>
</tr>
<tr>
<td>Guragone</td>
<td>pronoun is rarely used</td>
<td>(R. Green 1995: 130)</td>
</tr>
<tr>
<td>Bardi</td>
<td>pronoun is rare in speech, no anaphoric use</td>
<td>(Bowern 2004: 42)</td>
</tr>
<tr>
<td>Martuthunira</td>
<td>use of the pronoun is restricted</td>
<td>(Dench 1994: 100)</td>
</tr>
<tr>
<td>Wubuy</td>
<td>no reported restrictions</td>
<td>(Heath 1984: 241)</td>
</tr>
<tr>
<td>Arrernte</td>
<td>no reported restrictions</td>
<td>(Wilkins 1989: 124)</td>
</tr>
<tr>
<td>Yolngu</td>
<td>no reported restrictions</td>
<td>(Wilkinson 1991: 210)</td>
</tr>
<tr>
<td>Yuwaalaraay</td>
<td>no reported restrictions</td>
<td>(Giacon 2014)</td>
</tr>
</tbody>
</table>

Obviously, this does not mean that the lack of a third person pronoun would directly lead to the presence of anaphoric articles. However, the development of an anaphoric article (which always go back to an anaphoric pronominal form)\(^\text{16}\) 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 the language system may have also 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.

\(^\text{16}\)All anaphoric in the sample can also occur as pronominal forms. While I am not aware of any languages with an anaphoric article (or an admnominal 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).
8.3.2 Distribution across the referential space

The intuitive way to distinguish between different article types is to distinguish them according to their referential function, i.e. the (combinations of) referent types that they express. This was proposed in Chapter 3 based on the distinction of definite, specific, nonspecific, and generic referent types and led to the distinction of 8 major article types.

Another possible distinction of different article types can be made on the basis of the number of referents that the article encodes. We can conceptualize the three major referents type (definite, specific, and nonspecific) as the referential space that is relevant for articles. How many referents an article can mark thus corresponds to how much of the referential space the article covers. In this vein, Table 8.30 shows a typology of the articles that are attested in the world’s languages, distinguishing between the following three types: articles that encode a single referent type (in red), articles that encode two referent types (in blue), and the article encoding three referent types (in green). As was argued in the previous section, articles whose primary function it is to encode generic referents are not attested, which is why generic referents are not part of the referential space discussed here and are not represented in the referential space shown in Table 8.30.

Table 8.30: A typology of articles

<table>
<thead>
<tr>
<th>referent types</th>
<th>1 referent type</th>
<th>2 referent types</th>
<th>3 referent types</th>
</tr>
</thead>
<tbody>
<tr>
<td>specific</td>
<td>ART:EXSPEC</td>
<td></td>
<td>ART:REF</td>
</tr>
<tr>
<td>nonspecific</td>
<td>ART:NSPEC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As is shown in Table 8.30, there are three main articles types that belong to the first type (red) and that mark a single referent type: definite (DEF), anaphoric (ANA), and recognitionnal (RECOG) articles only encode definite referents; exclusive-specific articles (EXSPEC) encode specific referents, and nonspecific articles (NSPEC) mark nonspecific referents only. We find two article types in the world which co-express two referent types and therefore belong to the second type (blue): inclusive-specific articles (INSPEC) that do not distinguish between definite and specific referents, as well as indefinite (INDEF) articles that encode specific and nonspecific referents. The referential article (REF) corresponds to the third type (green), co-expressing definite, specific, and nonspecific referents. As was already mentioned in the previous sections, marking more than one referent type does not mean that the article is polysemous but that it is semantically vague with respect to the two functions coded.
Related to the typology of articles and the distribution of the referential space shown in Table 8.30, we can consider the number of articles in the sample by which each of the main referent type is expressed. This number follows from the distribution of article types in the sample presented in Table 8.31 below, repeating the counts for different article types given in Figure 8.14 (p. 226). Since we are interested in the number of articles that express the three major referent types in the sample, we can count the different types of definite articles together and distinguish between six relevant article types, as is shown in Table 8.31.

Table 8.31: Distribution of article types in the sample

<table>
<thead>
<tr>
<th>article type</th>
<th>N in sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite</td>
<td>82</td>
</tr>
<tr>
<td>of which anaphoric</td>
<td>27</td>
</tr>
<tr>
<td>of which recognitional</td>
<td>6</td>
</tr>
<tr>
<td>indefinite</td>
<td>36</td>
</tr>
<tr>
<td>exclusive-specific</td>
<td>12</td>
</tr>
<tr>
<td>nonspecific</td>
<td>5</td>
</tr>
<tr>
<td>inclusive-specific</td>
<td>7</td>
</tr>
<tr>
<td>referential</td>
<td>6</td>
</tr>
</tbody>
</table>

Following from the definitions of each of those article types, we can say that definite referents are expressed by definite, inclusive-specific, and referential articles. The number of articles in the sample that expresses definite referents thus consists of the sum of the articles of those three types, i.e. 82+7+6=95 articles. In the same vein, the number of articles expressing the specific referent type corresponds to the sum of inclusive-specific, exclusive-specific, and indefinite articles in the sample (49 articles). Finally, the number of articles in the sample that encode the nonspecific referent is 47, the sum of indefinite, nonspecific, and referential articles in the sample. Figure 8.21 illustrates those proportions by showing the number of articles (out of 148 articles in total) that are used to mark definite, specific, and nonspecific referents.
In Figure 8.21 we see the clear picture of definite referents being coded far more often in the sample than both specific and nonspecific referents. Table 8.32 summarizes the counts for each referent type being marked by an article in the sample together with the proportion.

Table 8.32: Observed proportions of the three referent types expressed by articles

<table>
<thead>
<tr>
<th>type expressed</th>
<th>N type expressed : N total expressed</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite</td>
<td>95 : 148</td>
<td>0.64</td>
</tr>
<tr>
<td>specific</td>
<td>49 : 148</td>
<td>0.33</td>
</tr>
<tr>
<td>nonspecific</td>
<td>47 : 148</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Fitting a binomial regression model confirms that there is an underlying difference between the numbers for encoded definite referents vs. specific and nonspecific referents, while the latter two referent types are underlyingly expressed by a similar number of articles. The model fitted is given in (43) and the estimated probability distributions for the proportions of the three referent types being expressed by an article are shown in Figure 8.21.

(43) A binomial regression that models the proportion of a referent expressed by an article as a function of the referent type. This estimates the association between the referent being expressed by an article and the type of referent.

\[(N \text{ type expressed} | \text{trials}(N \text{ total expressed}) - \text{referent type})\]
This is similar but not identical to the findings in Dryer (1989, 2013a,b) who observed that there are more definite than indefinite articles in the world’s languages. Also note that based on the sample of this study, we cannot be certain that the overall frequency of definite and indefinite articles differ in the world’s languages. The crucial difference between Dryer’s and the results of this study is that the proportions in Figure 8.21 do not reflect a claim about the frequency of article types in the world’s languages per se. The latter show the crosslinguistic trend of articles across the referential space i.e. which referent types are more likely to be coded by articles in the world’s languages. We can be certain that definite referents are significantly more likely to be encoded by an article than either specific or nonspecific referents, independently from concrete article types. From that follows that the impression of observing more definite articles than, for instance, indefinite articles in the world’s languages is rather a consequence of the general trend towards marking definite referents in the grammar than an independent trend itself. As I show in the following paragraphs, a typology of article systems allows us to find the same overall trend reflected in the frequency of certain types of article systems in the languages of the world: article systems that express definite referents differently from specific and nonspecific referents are crosslinguistically more frequent. Thus, the distribution of articles across languages reflect a more general trend that definite referents are generally more prominent in grammatical systems than both specific and nonspecific referents, given that they are more often coded by an article. This in turn justifies the general distinction between a definite and an indefinite domain.
8.3.3 Unattested article types

8.3.3.1 Generic articles

A generic referent was defined in section 3.2.4 as a referent that does not correspond to a single referent of its kind, but to all referents of its kind, i.e. to the kind set.

Articles whose primary function it is to mark generic referents do not seem to be attested in the languages of the world. This does not mean that articles cannot occur with generic referents; in many European languages, we find that generic referents are typically expressed by bare nouns in the plural or nouns in the singular together with a definite article. However, from a crosslinguistic point of view, we do not find a 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 (44) to (45) show two contexts with generic referents in Armenian. We see that the definite article is used together with a singular noun. Example (46) shows that Bullom So can also express generic referents as a plural noun together with the definite article.

(44) 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)

(45) mard-u-n kar-oł ē krakel-ov ē’-span-es bayc xosk’-ov
person-DAT-ART:DEF can-PTCP:PRS is shoot.INF-INST NEG-kill-FUT.2SG but word-INST
span-es kill-FUT.2SG
‘One cannot kill a human with shootings, but with words you kill him.’
Armenian (*Dum-Tragut 2009*: 162)

(46) [imúsmēnē əcē] ṇà cé-ēn āwōyēn
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, e.g. in Basque and Bemba. Examples (47) and (48) show the use of the inclusive specific article with generic referents in the singular and plural in Basque, and example (49) shows that Bemba can express a generic with the inclusive specific article and a plural noun as well. Baure, on the other hand, has a referential article. The latter can equally occur together with a noun whose referent is generic, which is shown in (50).
In other languages, the article types shown in the preceding examples are not compatible with generic referents. Both Sheko and Oko, for instance, have definite articles. Nouns that express generic referents, however, cannot be used with the definite article and occur as bare nouns. This is shown in (51) for Sheko and in (52) for Oko.

(51) gárgá íntʃu-rà há=gyá-mɔ
termite wood-ACC 3SG.M=chew-IRR
‘A termite eats wood.’ Sheko (Hellenthal 2010: 142)

(52) ógbén e-yíwo
child PROG-cry
‘A child is crying / Children are crying’ Oko (Atoyebi 2010: 161)

Bonan has an indefinite article. As is shown in (53) and (54), 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. As we see in example (55), generic referents are expressed by bare nouns, while the exclusive-specific article is not used.

(53) ɲantʃa=da ɭawa zaija bi-saŋ
past=LOC shaman strong COP.S-EPIST
‘In the past, shamans were very powerful.’ Bonan (Fried 2010: 83)

(54) talo ɲaŋda ʋɔsi ɬɔyɔsɔ ɬko ꜱwa
tall.building in dog raise difficult COP.O
‘Keeping dogs in an (apartment) building is difficult.’ Bonan (Fried 2010: 83)

(55) amzarái muɾ-u=bhaáu insaán na kha-áan-u
lion die.PTCP-MSG=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 with generic referents fundamentally differs from their uses in other contexts. Because articles of different types can occur with generic referents, we can assume that is not their primary referential function that makes them compatible with generic referents. 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.

It seems plausible that the article types that are attested with generic referents, i.e. definite, indefinite, specific, and referential articles, are articles with a relatively high token frequency in language use to begin with, which probably allowed them to extend their use to contexts with generic referents. Minor article types\textsuperscript{17} such as anaphoric articles, recognition articles, or presentation articles cannot be used. One possible explanation of this distinct behaviour of different article types with respect to their compatibility with generic referents is based on their functional domains, meaning that semantically broader articles may have more potential to extend their functional domain than articles with a more restricted meaning. Another way to account for this difference with respect to the use of articles with generic referents is based on the distribution of articles in the language. If a language encodes several referent types by articles and if this leads to bare nouns being infrequent in the language, the system could allow or require the use of articles with generic referents in order to avoid the occurrence of bare nouns.

Across languages, the high variation that we find with respect to articles and their use with generic referents in addition to the lack of evidence for articles whose primary function it is to encode generic referents lead to the conclusion that generic articles do not exist. That generic referents do not have a marker of their own is plausible from a functional perspective. The generic referent is the least individuated referent type, and arguably the least prominent referent type in our communicative needs, which would also make it the least frequent referent type in language use. Therefore, it is not surprising that languages do not develop a grammatical marker, i.e. an article, whose main function is the expression of generic referents.

8.3.3.2 Non-anaphoric definite articles

I proposed three types of articles that belong to the definite domain in this study: definite, anaphoric, and recognition articles. Previous studies suggested the distinction of another type of definite articles, namely non-anaphoric definite articles based on the data of a handful of languages with markers which seem like definite articles, but which are not used to encode anaphoric referents. In this section, I discuss the data that has led to the proposal of this article type and I argue that the

\textsuperscript{17}Minor is understood here in the sense of a narrower function of the article, which leads to its occurrence in fewer contexts and makes it less frequent in language use.
evidence for non-anaphoric definite articles is not strong, which is why non-anaphoric definite articles were not included as a major article type in this study.

In his reference hierarchy, Dryer (2014) predicts the existence of non-anaphoric definite articles, mentioning the language Ma’di as having such an article. Also Schwarz (2009, 2013) provides a detailed discussion of two distinct types of definite articles co-occurring in individual languages, one of which is the non-anaphoric definite article. He labels those “weak” definite articles and opposes them to “strong” definite articles which largely correspond to what I call anaphoric articles.

Definite articles that cannot encode an anaphoric referent are very rare in the world’s languages and it seems that there are only two types of scenarios that lead to their development. The first scenario involves an emerging definite article that does not originate from a demonstrative, but, e.g., from a possessive. In its first stages, the marker can extend to the functions of a definite article without necessarily extending its use to anaphoric contexts “immediately”. However, many languages with a definite article originating from a possessive marker (e.g. Sumu, Amharic) show that eventually, the article is very likely to extend its use to anaphoric contexts as well. An example of an emerging definite article that has not yet entirely reached this stage is Indonesian.

In Indonesian, as was mentioned in section 7.1, the third person possessive marker -nya (which is also used as an object pronoun) can be used to mark referents as definite based on situational uniqueness, which can be viewed as an extension of the original possessive semantics of the marker and as an emerging definite article. While it is still usually the demonstrative that is used to mark anaphoric referents, Rubin (2010: 109) notes that in colloquial speech, -nya can be used instead. Examples (56) and (57), repeated from section 7.1, show -nya in a uniqueness-based and in an anaphoric context, respectively.

(56) kalau mau makan nasi-nya di lemari
    if want eat rice-DEF in pantry
    ‘If you want to eat, the rice is in the pantry.’ Indonesian (Rubin 2010: 107)

(57) 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 deket rumah?
    salon-DEF in near house
    ‘Is the salon near your house?’ Indonesian (Rubin 2010: 109)

The second scenario that may give rise to a non-anaphoric definite article includes an article system with a definite article and an additional anaphoric article. If the anaphoric article can block the use of the broader definite article to encode anaphoric referents, the definite article corresponds to a non-anaphoric definite article. The languages that were mentioned to have such
a system are Ma’di (Dryer 2014) and Lakota as well as a number of Germanic languages (Schwarz 2013). My sample also contains one language, Jamsay, with both an anaphoric article and a definite article, making the latter a candidate of a non-anaphoric definite article. I briefly present the definite articles from these languages, showing that most of them cannot be regarded as proper non-anaphoric definite article; since in most cases, the presence of the anaphoric articles does not automatically block the use of the definite article in anaphoric contexts.

In Jamsay and Ma’di, even though they have a dedicated anaphoric article, the latter does not automatically block the use of the definite article with anaphoric referents. Heath (2008: 164) notes for Jamsay that the anaphoric article kò is often used together with the definite article kùⁿ. This is illustrated in (26) below.

(58) tògù pó:ró tógó kó b̀èrè: bè nû: [kò tòg kùⁿ] úró
shed first shed building 3SG.NHUM in 3PL.enter PfV ART:ANA shed ART:DEF house
tána-ŋá mèyⁿ nínŋ y7=k3
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)

For Ma’di, Blackings & Fabb (2003) distinguish between a definite article rì and an anaphoric article 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)” (Blackings & Fabb 2003: 17). This description also suggests that even though Ma’di has an anaphoric article, the definite article is not excluded from marking anaphoric referents. What complicates the picture of the two articles in Ma’di is the status of the definite article. Throughout the grammar, the so-called definite article often occurs together with the focus marker ʔɪ̄ and also its translation as “the … in question” make it questionable whether the marker is a definite article at all. Without discussing its status here, it can be shown that it is not excluded from occurring with anaphoric referents. Example (59a) shows an utterance referring to ɓá ‘people’. Two sentences later in the conversation, (59b) is uttered, referring back to the referent of ɓá ‘people’. We see that the noun is marked by the so-called definite article rì and not by the anaphoric article nā.

(59) a. ādrúpì ní tfjáragùlè k-ɓù ʔání ní ɓá mià áziá
brother 3SG Caragule 3-call 3S BEN people hundred six
‘His brother Caragule should get the service of six hundred people for him.’

b. îtó ʔà ādrúpì ūngwē [ɓá rì] ʔì
Ito his brother call people ART:DEF FOC
‘Hare’s brother called them, the people.’

Ma’di (Blackings & Fabb 2003: 674)
Assuming that $i$ in Ma’di is a definite article, its relation to the anaphoric article is not essentially different from the relation between definite articles and demonstratives in the languages in which the definite article can also occur to mark referents as deictic and thus competes with the demonstrative in deictic and also anaphoric contexts.

Lakota is another language with similar properties in that it was reported to have an anaphoric and a definite article, which is why Schwarz (2013) claims that the definite article in Lakota is restricted to the marking of only uniqueness-based referents, similarly to what was claimed for Ma’di by Dryer (2014). While it could be shown for Ma’di and Jamsay that the presence of an anaphoric article does not automatically block the definite article in anaphoric contexts, the situation in Lakota is different.

The use of the definite article $ki$ in Lakota being well attested (Curl 1999, Rood & Taylor 1996, Van Valin 1977, Williamson 1984), the so-called anaphoric article $kuŋ$ is also mentioned in the works cited above, but not documented by convincing examples in any of these studies, since the scarce evidence shown always goes back to Buechel (1939) with respect to the anaphoric article $kuŋ$. Also in Schwarz’s discussion on the two articles in Lakota, the only source for examples of $kuŋ$ is Buechel (1939). Moreover, Curl (1999) and Ingham (2003) note that $ki$ is the regular definite article which can also be used anaphorically, and even though $kuŋ$ often marks anaphoric referents, its main function is the indication of a topic-shift. Therefore, I conclude that $kuŋ$ in Lakota cannot be regarded as an anaphoric article. Thus, Lakota is not a convincing example of a language with a definite non-anaphoric article either.

Schwarz (2009) also discusses data from several West Germanic languages, one of which is Fering (North Frisian, Germany). In Fering, it seems to be the case that that the existence of an anaphoric article blocks the use of the definite article in anaphoric contexts. Example (60) below shows a minimal pair, contrasting a situationally unique (60a) with an anaphoric (60b) referent.

We see that the two articles, definite $a$ and anaphoric $di$ are mutually exclusive in these contexts.

(60) a. ik skal deel tu [a/ *di] kuupmaan
   I must down to ART:DEF ART:ANA grocer
   ‘I have to go to the grocer.’

   b. oki hee an hingst keeft. [di/ *a hingst] haaltet
   Oki has a horse bought ART:ANA ART:DEF horse limps
   ‘Oki has bought a horse. The horse limps.’

   Fering (Schwarz (2013: 538), cited from Ebert (1971)

The same effect can be found in German with definite articles following certain prepositions. Together with these prepositions, the singular forms of the definite article occur in a contracted

---

18Lakota is also part of my sample; I do not treat $kuŋ$ as an article for the reasons mentioned above.
form which can only be used in non-anaphoric definite contexts. At the same time, the use of the uncontracted combination of the preposition and the definite article is not acceptable in such contexts. An example to illustrate this is given in (61). Example (62) shows that a situationally unique definite referent, on the other hand, requires the contracted form, if available:

(61) In der New Yorker Bibliothek gibt es ein Buch über Topinambur. Neulich war ich dort und habe [in dem /#im Buch] ... ‘In the New York public library, there is a book about topinambur. Recently, I was there and (searched) the book (for an answer to the question of whether one can grill topinambur).’

German (Schwarz 2009: 30)

(62) Der Empfang wurde [vom /#von dem Bürgermeister] eröffnet. ‘The reception was opened by the mayor.’

German (Schwarz 2009: 40)

I did not consider this distinction between a definite and an anaphoric article in German in the other parts of this study because it is lexically restricted by the preposition used.

The data from Fering and German are similar to examples (37) and (38) from Kaqchikel showed in section 4.1.3, repeated in (63) and (64) below.

(63) a Lu xiroyoj jun b’eychik mambajota yich’o ruk’in [ri vinaq *(ri)]
    PROG Pedro call.PST.S:3SG.O:1SG one way want.NEG.1SG talk with.3SG ANA person

    ‘Pedro called again. I don’t want to talk to that person.’

Kaqchikel (prim. data)

(64) k’ab’a a Lu xuroyoj ri ru vecino loman chuk’a 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 (prim. data)

Although it was shown in section 4.1.1 that the definite article in Kaqchikel occurs in anaphoric contexts as well, examples like (63) and (64) show that its use can be blocked in certain types of anaphoric contexts. The most influential factor probably is the distance between the antecedent and the anaphoric referent. What these examples hint at is that even though the use of the definite article can be blocked with certain types of anaphoric referents by an anaphoric article, the restriction depends on certain properties of the anaphoric referent, and do not apply to all anaphoric
contexts. Therefore, and also because the number of reported non-anaphoric articles is so low, it is questionable if non-anaphoric definite articles in the strict sense exist at all in the world’s languages.

To conclude, non-anaphoric definite articles are extremely rare across the languages of the world, and non-anaphoric definite articles in the strict sense probably correspond to the idealization of a more complex competition pattern between two articles with respect to different types of anaphoric referents.

8.4 Article systems

In the first part (section 8.4.1), this section offers a typology of the coverage of the referential space by the entire article system. In addition to systematizing the attested article systems in the languages of the world, this typology is useful to address the question of systematic crosslinguistic gaps. Section 8.4.2 turns to article systems that do not entirely fit into the typology discussed in section 8.4.1, because they feature articles that overlap in their functions. As will be shown, different kinds of functional overlap within article systems can be accounted for by general discourse-pragmatic principles.

8.4.1 A typology of article systems

Similarly to the typology of articles presented in Table 8.30 in the previous section, article system of individual languages can be distinguished on the basis of the number of referents that the article system encodes. In other words, we can describe article systems in terms of the referential space that they cover. Again, I will consider anaphoric and recognitional articles together with definite articles for the purposes of this section. We will see in this section that the major types of attested article systems in the world’s languages can be divided into systems that encode one, two, or three referents types.

8.4.1.1 Article systems that cover the referential space minimally

Article systems that encode one referent type can logically only consist of a single article. The only article types that are attested as single articles in the language are definite, including, anaphoric, and recognitional, articles, as well as exclusive specific articles. Those two systems are presented in Table 8.33 as system I and system II, respectively. I am not aware of any language...
with a nonspecific article that would occur without other articles within that language (system ‘I). Table 8.34 lists the languages in the sample that have those two article types from this series.

<table>
<thead>
<tr>
<th>referent type</th>
<th>system I</th>
<th>system II</th>
<th>system ‘I</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite</td>
<td>ART:DEF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>specific</td>
<td>ART:EXSPEC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nonspecific</td>
<td></td>
<td>ART:NSPEC</td>
<td></td>
</tr>
</tbody>
</table>

Table 8.34: Distribution of article systems in the sample (first series)

<table>
<thead>
<tr>
<th>article system</th>
<th>languages in sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>system I</td>
<td>N = 54</td>
</tr>
<tr>
<td>system II</td>
<td>N = 3</td>
</tr>
<tr>
<td>Angolar, Barwar, Rajbanshi</td>
<td></td>
</tr>
</tbody>
</table>

System I, consisting of a definite article only, is by far the most frequent one and occurs in 54 languages of the sample, while language with an exclusive-specific article are less frequent; 3 languages of the sample have a system of type II. An example of type I is Mokpe, whose definite article was discussed in detail in section 4.1.2. The following examples, repeated from section 4.1.2, illustrate how an article system of type I covers the referential space: only definite referents are marked by an article (65), leaving both specific (66) and indefinite referents (67) unmarked i.e. expressed as a bare noun.

(65) [*e) mòkanëli] à 3óβ-è bòndà jà [*e) ndʒumá]
ART:DEF leader.C1 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 (prim. data)
(66) nò óbélì lómbá
 1sg.2sg have bundle
‘I have a gift for you.’ Mokpe (prim. data)

(67) ówélì èkì já lì lá èné mbówá
  there.is place of INF eat dem:prox village.Q
‘Is there a place to eat in this village?’ Mokpe (prim. data)

An example of system II with an exclusive-specific article only can be found in Angolar (Creole, Sao Tome and Principe). The only article in the language, uⁿa, is used to mark specific referents (68), while both definite (69) and nonspecific referents (70) are expressed by a bare noun.

(68) Nda ma vutuka fo miônga 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)

(69) N ka tango [uⁿa ोoya].
 1sg fut tell ART:EXSPEC story
‘I will tell a story.’ Angolar (Lorenzino 1998: 133)

(70) Ši n ka vuna kikie ngai n na ta ma n ka паθa 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)

8.4.1.2 Article systems that cover most of the referential space

Article systems that encode two referent types are shown in Table 8.35. One way to group these systems is to consider the referential space they cover as a whole. Systems III and IV (in blue) cover the definite and the specific domain, leaving nonspecific referents unmarked in the language, while systems V and VI (in red) cover the indefinite domain but not definite referents. Another way to group the systems that encode 2 referent types is to look at the number of articles present. There can either be a single article that co-expresses two referent types: the inclusive-specific article (system IV) and the indefinite article (system VI). Or, the system consists of two separate articles for each referent type: system III, consisting of a definite article as well as system V with an exclusive-specific article and a nonspecific article. The combination of an exclusive specific article within a nonspecific article (system V) is not attested in the sample, as can be seen in Table 8.36. However, as was shown in Chapter 5, Q’anjobal may be considered as an example of a language with a specific and nonspecific article and with no definite article. The two systems *II and *III that are logically possible but not attested both feature an article expressing nonspecific referents in the absence of an article that would express a specific referent. System *II consists of two separate articles for definite and nonspecific referents, while system *III contains a single,
hypothetical definite-nonspecific article that would co-express definite and nonspecific referents (but crucially not specific referents). System *III points towards the hierarchical structure between definite, specific, and nonspecific referents: it is another piece of evidence supporting the referential hierarchy that is also found in other parts of grammar.

Table 8.35: Article systems encoding 2 referent types

<table>
<thead>
<tr>
<th>referent type</th>
<th>system III</th>
<th>system IV</th>
<th>system V</th>
<th>system VI</th>
<th>system *II</th>
<th>system *III</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite</td>
<td>ART:DEF</td>
<td>ART:INSPEC</td>
<td></td>
<td></td>
<td>ART:DEF</td>
<td>ART:DEFNSPEC</td>
</tr>
<tr>
<td>nonspecific</td>
<td>ART:NSPEC</td>
<td></td>
<td>ART:NSPEC</td>
<td></td>
<td>ART:NSPEC</td>
<td>ART:DEFNSPEC</td>
</tr>
</tbody>
</table>

Table 8.36: Distribution of article systems in the sample (second series)

<table>
<thead>
<tr>
<th>article system</th>
<th>languages in sample</th>
</tr>
</thead>
</table>
| system III     | N = 8
Agta, Akan, Buwal, Komnzo, Konso, Logba, Oko, Palula |
| system IV      | N = 1
Runyankore |
| system V       | N = 0 |
| system VI      | N = 20
Aghul, Asheninka Perené, Bilua, Bonan, Cape Verdean Creole, Carib, Chatino, Chocholtec, Domari, Kashibo-Kakataibo, Kharia, Kurtöp, Lango, Mapudungun, Nhengatu, Parecis, Sri Lanka Malay, Sye, Tamil, Turkish |

System III with a definite article and an exclusive-specific article is found in 8 languages of the sample, one of which being Akan whose anaphoric and exclusive-specific articles were discussed in sections 4.2.2 and 5.1.2, respectively. The following examples illustrate this article system with data from Buwal (Chadic, Cameroon): we see how definite (71) and specific (72) referents are marked by two separate articles, while nonspecific referents (73) are expressed as a bare noun.
(71) Hwa-baw taf ma=kəɗa uza, [taf anta] a-baw, a-nda ta ƞ s:2sg-turn path rel=towards down [path ART:DEF] s:3sg-turn s:3sg-go through prep lekwal.
school
‘You turn on the path which goes down, the path turns, it goes through the school.’

Buwal (Viljoen 2013: 239)

(72) [Ŋhwə-ye vedaye] a wata juraw ete a Zukwadʃaŋw aka.
goat−pl ART:EXSPEC.PL PREP compound sub.chief here PREP Zukodfong exist
‘There are some goats at the sub-chief’s home here in Zukodfong.’

Buwal (Viljoen 2013: 455)

(73) Berjeŋ naka akwaw.
donkey POSS:1SG NEG.EXIST
‘I don’t have a donkey.’

Buwal (Viljoen 2013: 490)

Languages with an article system of type IV, which consists of a single article covering both definite and specific referents, are typologically extremely rare as well and do not seem stable. The one languages in the sample that is classified to consist of an inclusive-specific article only is Runyankore (Bantu, Uganda). However, Runyankore actually has two articles, the other one being an anaphoric article which does not express any additional referent type so that the system is counted as type IV here (but cf. the discussion in section 8.4.1.4).

A probably better example of an article system of type IV is found in Bemba (Bantu, Zambia).20

As was discussed in section 6.1.1, Bemba has an inclusive-specific article that is used systematically in both definite and specific contexts, while nonspecific referents only receive a class prefix but not the article prefix that precedes the latter. Example (74), repeated from section 6.1.1 show how the use of the article cannot distinguish between a definite and a specific interpretation of the referent. In (75), on the other hand, we see a nonspecific referent with no article present.

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

(75) nshiamwene mu-ana
see.PST.NEG.1SG CL1-child
‘I didn’t see any child.’
Bemba (Givón 1969: 42)

A similar situation holds for the next type of article systems, type V, consisting of an exclusive-specific article and a nonspecific article and without a marker that would code definite referents. As was mentioned in the beginning of this section, this type of system is not attested in the

20Bemba is not part of the sample, which is why it does not appear in any counts.

250
sample, but Q’anjobal may be a candidate with an exclusive-specific and a nonspecific article, but no definite one. Examples (76) to (78) illustrate this: In (77) and (78), we see the exclusive-specific article *jun* and the nonspecific article *junoq* with a specific and a nonspecific referent, respectively. Example (76) features a definite referent. While it is not marked by an article as such, definite referents are often not expressed as bare nouns in Q’anjobal either. As we see in (76), the definite referent of *anima* ‘person’ is accompanied by the classifier *cham*.

(76) *cham* *anima* may kon tojlaneni ay *jun* miman ya-tut.
    *cl:*male person *art:*exspec pay.*pst:*3sg.o:*1pl exist big 3sg-house
    ‘The man who paid us has a big house.’

(77) *tzeb’ach* yul *jun* tuktuk. mayal wawrtej naq tz’umon ch’en
    come.imp in *art:*exspec tuktuk already called *cl:*driver
    ‘Let’s take a (certain) tuktuk. I already called the driver.’

(78) *asi’* yul *junoq* tuktuk.
    go.imp in *art:*nspec tuktuk
    ‘Let’s take a (any) tuktuk.’

Q’anjobal uses 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’anjobal do not indicate the referential status of the noun. The presence of the classifier with definite referents may however be relevant to the article system insofar that it could negatively influence the development of a definite article, given that definite nouns do not necessarily occur as bare nouns as opposed to specific and nonspecific referents.

The last system of this series of article systems that cover two of the three referent types consists of a single indefinite article. In contrast to the previous one, this system is relatively common and occurs in 20 languages in the sample. The examples below illustrate system VI with data from Kashibo-Kakataibo (Panoan, Peru). That definite referents are expressed by a bare noun is shown in example (79b): the referent of *gringo* is anaphoric in and occurs without an article. In (80) and (81), on the other hand, we see that the same article, *achusi*, occurs with a specific and a nonspecific referent, respectively. As *achusi* is restricted to discourse-prominent referents, it is glossed as a presentational article, which corresponds to a subtype of indefinite articles.

(79) a. *kaisa* is is-ishi-akë-x-a atu-n.
    3 see see-only-*rempst:*3-dist *a:*3pl
    ‘They only looked at him several times.’

251
b. *atian gringo a-n=ribi kaisa is-ishi-akë-x-in atu a xubu=nu then **white person** 3sg-a=also 3 see-only-rempst-3-prox they that house=loc kaisa nukut-akë-x-in achushi xubu=nu atu=n xubu=nu.

3 arrive-rempst-3-prox art:pres house=loc 3pl=gen house=loc

‘Then, the gringo also used to look at them, and arrived at their houses.’

Kashibo-Kakataibo (Zariquiey Biondi 2011: 757)

(80) a=x ka ‘ikën [achusi matá] ka is!

that=3:narr be.3 art:pres hill.abs narr look

‘There is a hill, look!’

Kashibo-Kakataibo (Zariquiey Biondi 2011: 250)

(81) ’ë=x Lima=nu kwan-xun kana [achushi casaca] bits-kê ’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)

8.4.1.3 Article systems that cover the entire referential space

The last series of article systems is shown in Table 8.37; it consists of article systems that mark all three referent types, i.e. cover the entire referential space relevant to article systems. Such systems can consist of two articles (in blue), of three separate articles (in red), or of a single article (in green) used for all three referent types. The two attested systems with two articles in this series are the following: system VII with a definite and an indefinite article as well as system VIII with an inclusive-specific article and a nonspecific article. System IX encodes each referent type by a separate article, namely a definite, an exclusive-specific, and a nonspecific article. The last attested type of article systems in this series, system X, consists of a single referential article, covering the entire referential space. The unattested system that would cover the referential space, system *IV, corresponds to system *III with an additional exclusive-specific article. This typological gap, too, can be captured by the hierarchical relation between definite, specific, and nonspecific referent types: an article (or any marker, for that matter) cannot co-express definite and nonspecific referents without expressing specific referents. Table 8.38 shows the languages in the sample that feature an article system of this series.

<table>
<thead>
<tr>
<th>referent type</th>
<th>system VII</th>
<th>system VIII</th>
<th>system XI</th>
<th>system X</th>
<th>system *IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite</td>
<td>art:def</td>
<td>art:defnspec</td>
<td>art:def</td>
<td>art:ref</td>
<td>art:defnspec</td>
</tr>
<tr>
<td>specific</td>
<td>art:inspec</td>
<td>art:exspec</td>
<td>art:exspec</td>
<td>art:ref</td>
<td>art:exspec</td>
</tr>
<tr>
<td>nonspecific</td>
<td>art:indef</td>
<td>art:nspec</td>
<td>art:nspec</td>
<td>art:ref</td>
<td>art:exspec</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8.38: Distribution of article systems in the sample (third series)

<table>
<thead>
<tr>
<th>article system</th>
<th>languages in sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>system VII</td>
<td>N = 13</td>
</tr>
<tr>
<td></td>
<td>Albanian, Arawak, Armenian, Chol, Chumash, German, Hungarian, Lavukaleve, Oksapmin, Pipil, Teiwa, Ulwa, Wayuu</td>
</tr>
<tr>
<td>system VIII</td>
<td>N = 3</td>
</tr>
<tr>
<td></td>
<td>Ayoreo, Siar Lak, Tongan</td>
</tr>
<tr>
<td>system XI</td>
<td>N = 2</td>
</tr>
<tr>
<td></td>
<td>Biak, Lakota</td>
</tr>
<tr>
<td>system X</td>
<td>N = 6</td>
</tr>
<tr>
<td></td>
<td>Baure, Halkomelem, Mamaindê, Rapa Nui, Sabanê, Sundwadia</td>
</tr>
</tbody>
</table>

System VII is not only what we find in many European languages, it is also the system that is the most frequent one of this series, occurring in 13 languages in the sample. This article system is however not restricted to Europe; the following examples show for Pipil (Aztec, El Salvador) that it uses a definite article with definite referents (82), and an indefinite article for both specific (83) and nonspecific (84) referents.

(82) entonces [ne i-siwa:-w] ki-mik-tih ne chumpipi, ki-chiw-ki desplumár, wan ki-chiw-ki ne komidah wan ki-kwah-ke-t. pluck and 3sg-do-pst ART:DEF food and 3sg-eat-pst-pl

‘Then, his wife killed the turkey, plugged it, and prepared the food and they ate it.’ (Campbell 1985: 868)


‘In those times hurricaners came from a town named Chiltiupan.’ (Campbell 1985: 867)

(84) ni-mits-maka-skiya [se: mu-tamal], pero tesu ni-k-piya.

‘I would give you a tortilla, but I don’t have any.’ (Campbell 1985: 123)

The next system of this series, system VIII, consists of an inclusive-specific and a nonspecific article. In contrast to its 2-article counterpart with a definite and an indefinite article, this system is rare across languages and only attested in 3 languages of the sample. One of the languages is Ayoreo (Zamuco-Ayoreo, Bolivia, Paraguay). Ayoreo nouns have gender and number and both
articles inflect for the two categories; they occur as suffixes on the noun. Table 8.39 shows the two paradigms of the inclusive-specific and the nonspecific article in Ayoreo.

### Table 8.39: Article exponents in Ayoreo (Ciucci 2016: 456-462, 475-481)

<table>
<thead>
<tr>
<th>gender</th>
<th>number</th>
<th>inclusive-specific</th>
<th>nonspecific</th>
</tr>
</thead>
<tbody>
<tr>
<td>feminine</td>
<td>singular</td>
<td>-i/-u/-ia/-∅</td>
<td>-rak/-tak/-nak</td>
</tr>
<tr>
<td></td>
<td>plural</td>
<td>-die/-i</td>
<td>-rigi/-tigi/-niŋi</td>
</tr>
<tr>
<td>masculine</td>
<td>singular</td>
<td>-i</td>
<td>-tik/-rik/-nik</td>
</tr>
<tr>
<td></td>
<td>plural</td>
<td>-ode</td>
<td>-tigo/-riŋo/-niŋo</td>
</tr>
</tbody>
</table>

Examples (85) and (86) show the use of the inclusive-specific article in definite and specific contexts, respectively. Example (87) in contrast illustrates the use of the nonspecific article with a nonspecific referent.


‘But in order for them to put down their own rifles, they should not have fear.’

(Bertinetto 2009: 54)


‘In Tumuchucua there is a hangar for the airplanes.’

(Ciucci 2016: 486)

(87) Mu que ore i-plata-rigi cuse.

‘But they have no money.’

(Ciucci 2016: 356)

Article systems of type IX, with a definite, exclusive-specific, and non-specific article are crosslinguistically very rare as well, and are attested in 2 languages in the sample, namely in Biak (Austronesian, Indonesia) and Lakota (Siouan, USA). Another example of a related Siouan language with such an article system is Crow. In Crow, articles appear as suffixes on the noun as well. We see in examples (88) to (90) that definite, specific, and nonspecific contexts require a separate article each.

(88) káalee-sh dàawi-kawe-h he-m old.woman-ART:DEF go.on-polit-imp say-ds

‘The old woman said, “go on”.

(Crow (Graczyk 2007: 239)
(89) \textit{dakáak-kaata-} \textit{m} \textit{húu-laa} \textit{hii-ka}  \\
\textit{bird-dim-art:inspec} come-ss reach-decl  \\
‘A bird came, it reached him.’  \\
\textit{Crow (Graczyk 2007: 228)}

(90) \textit{axée} \textit{baláxxiikaashe dúup-} \textit{eem} \textit{alúutkaashe áappaa} \textit{día-a-wa-ku-he}  \\
\textit{father bow} \textit{two-art:nspec} \textit{arrow} along \textit{with make-cont-1-give-aff-q}  \\
‘Father, will you please make me two bows and arrows?’  \\
\textit{Crow (Graczyk 2007: 230)}

The last system of this series, covering the entire referential space relevant to articles is system X, consisting of a single referential article. Such a system occurs in 6 languages in the sample; it is not very common crosslinguistically but it seems less rare than the two previous systems discussed. This article type was extensively discussed with data from Rapa Nui, Halkomelem, and Baure in section 6.2; examples (91) to (93) illustrate for Rapa Nui how the same article \textit{te} is used with all three major referent types.

(91) \textit{I} \textit{hāŋai era i nunui era, te aŋa o tú rū’au māmā era he}  \\
\textit{pfv} \textit{raise} \textit{dist pfv pl.big dist art:ref} \textit{work of dem old.woman mother dist pred}  \\
kā i \textit{[te} \textit{'umu]} \textit{paurō te mahana}.  \\
\textit{kindle acc art:ref} \textit{earth.oven every art:ref} \textit{day}  \\
‘When they had raised them and they had grown up, what the old mother did was cooking food in the earth oven every day.’  \\
\textit{(Kieviet 2017: 579-580)}

(92) \textit{'i te} \textit{noho iŋa tuai era} \textit{ā [te taŋata e tahi] te } \textit{’īŋoa}  \\
\textit{at art:ref} \textit{stay nmlz ancient dist ident art:ref} \textit{man num one art:ref} \textit{name}  \\
\textit{ko Tu’uhakararo} \textit{prom Tu’uhakararo}  \\
‘In the old times (there was) a man called Tu’uhakararo.’  \\
\textit{(Kieviet 2017: 238)}

(93) \textit{...mo ai o te} \textit{moni mo ho’o mai i [te haraoa].}  \\
\textit{...for exist of art:ref} \textit{money for trade hither acc art:ref} \textit{bread}  \\
‘(He sells food) in order to have money to buy bread.’  \\
\textit{(Kieviet 2017: 238)}

8.4.1.4 The crosslinguistic distribution of article systems

This section examines the crosslinguistic distributions of the different article systems presented in the previous section. As was mentioned, the minor article types from the definite domain were treated as definite articles for the purposes of this typology. For instance, Limbum has a single anaphoric article and the language was classified as having an article system of type I, corresponding to a single definite article. Similarly, Palula, with an anaphoric and an exclusive-specific article, was counted as having the system of type III, a definite and an exclusive-specific article. Languages that feature two separate articles in the definite domain, i.e. a definite with an
anaphoric or recognitional article were treated as having one definite article. The motivation for this choice is that all three article types (definite, anaphoric, recognitional) code the same referent type, namely definite referents or a more restricted subtype thereof. Applying this to the data set, Urama with a definite and an anaphoric article was classified as having a system consisting of a definite article. Mutatis mutandis, the same explanation holds for the classification of Runyankore (anaphoric and inclusive-specific articles) as a system with an inclusive-specific article.

In addition to this kind of functional overlap between articles within languages, 3 languages in the sample (Tepehua, Basque, Maori) are excluded from this typology, because they have article systems that show overlaps that make them substantially differ from the article systems distinguished in the previous sections. Addressing the issues of both types of functional overlap, section 8.4.2 discusses the article systems of those languages.

A more conservative treatment of article systems would exclude systems with recognitional and anaphoric articles, for both those that occur as single article in the definite domain and those that occur along another article covering definite referents in a given language. Excluding those systems from the sample leaves us with the reduced sample shown in Table 8.40. In the remainder of this section, the cross-linguistic distributions of article systems in the reduced sample will be discussed alongside with the ones of the full sample. As will be shown, the main trends are not affected by the choice of including or excluding article systems with the minor article types (anaphoric and recognitional).
Table 8.40: Distribution of article systems in the reduced sample

<table>
<thead>
<tr>
<th>article system</th>
<th>languages in the reduced sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>system I</td>
<td>N= 28</td>
</tr>
<tr>
<td>DEF</td>
<td>Akuntsú, Apinayé, Bajau, Balanta Ganja, Bambassi, Beng, Bullom So, Chimariko, Cupeno, Dime, Diyari, Gaahmg, Haida, Hakhun Tangsa, Irish, Koyra Chiini, Lepcha, Menya, Mokpe, Nuuchahnulth, Papuan Malay, Sandawe, Sheko, Supyire, Tarahumara, Ute, Yurok, Zoque</td>
</tr>
<tr>
<td>system II</td>
<td>N = 3</td>
</tr>
<tr>
<td>EXSPEC</td>
<td>Angolar, Barwar, Rajbanshi</td>
</tr>
<tr>
<td>system III</td>
<td>N = 5</td>
</tr>
<tr>
<td>DEF + EXSPEC</td>
<td>Agta, Buwal, Konso, Logba, Oko</td>
</tr>
<tr>
<td>system IV</td>
<td>N = 0</td>
</tr>
<tr>
<td>INSPEC</td>
<td></td>
</tr>
<tr>
<td>system V</td>
<td>N = 0</td>
</tr>
<tr>
<td>EXSPEC + NSPEC</td>
<td></td>
</tr>
<tr>
<td>system VI</td>
<td>N = 20</td>
</tr>
<tr>
<td>INDEF</td>
<td>Aghul, Asheninka Perené, Bilua, Bonan, Cape Verdean Creole, Carib, Chatino, Chochole, Domari, Kashibo-Kakataibo, Kharia, Kurtöp, Lango, Mapudungun, Nhengatu, Parecis, Sri Lanka Malay, Sye, Tamil, Turkish</td>
</tr>
<tr>
<td>system VII</td>
<td>N = 10</td>
</tr>
<tr>
<td>DEF + INDEF</td>
<td>Albanian, Arawak, Armenian, Chol, Chumash, German, Hungarian, Pipil, Ulwa, Wayuu</td>
</tr>
<tr>
<td>system VIII</td>
<td>N = 3</td>
</tr>
<tr>
<td>INSPEC + NSPEC</td>
<td>Ayoreo, Siar Lak, Tongan</td>
</tr>
<tr>
<td>system XI</td>
<td>N = 2</td>
</tr>
<tr>
<td>DEF + EXSPEC +NSPEC</td>
<td>Biak, Lakota</td>
</tr>
<tr>
<td>system X</td>
<td>N = 6</td>
</tr>
<tr>
<td>REF</td>
<td>Baure, Halkomelem, Mamaindê, Rapa Nui, Sabanê, Sundwadia</td>
</tr>
</tbody>
</table>

Figure 8.22 shows the distribution of the 10 different article systems across the six macro areas in the full sample on the left and in the reduced sample on the right. The more conservative reduced sample shows a similar general distribution of article systems as the full sample; the important difference is that the reduced sample, based on the areal distribution of anaphoric and recognitional articles, mainly affects Australia and Papunesia. What Figure 8.22 also shows is that there is much more areal variation than trends with regard to the distribution of different article sys-
tems. The main trends that we see follow from the distributions of article types: Australia, having
definite and anaphoric articles only, features systems of type 1 exclusively. The relatively high
proportion of system 1 in Africa and Australia is also partly a consequence of the high frequency
of definite article types. However, Figure 8.22 shows an areal trend for the three complex article
systems of the third series that are very rare crosslinguistically: system VIII (inclusive-specific
and non-specific) and system IX (definite, exclusive-specific, nonspecific) and system X (referential
article). These three types, marked in green and blue in Figure 8.22 only occur in the Americas
and in Papunesia.

Thus, while it can be shown to a certain degree that the distribution of article systems follows the
areal distribution of single article types, Figure 8.22 shows that there is a high degree of variation
and there seems to be no straightforward principle that would capture the variation that we find.
Another look at the distribution of the different article systems reveals that their crosslinguistic
frequency can in fact be related to a more general common principle. Figure 8.23 shows the
frequency distribution of different article systems according to the series that they belong to: the
systems that encode one referent type (in orange), the systems that encode two referent types
(in light yellow), and the systems that encode all three referent types (in green). The left plot

---

Figure 8.22: Distribution of articles systems across macro areas
shows this for the full data set and the right plot for the reduced data set excluding anaphoric and recognitional articles.

Both plots in Figure 8.23 show that for each series, there is a single system that occurs more frequently than the other system(s) of the same series. For systems of the first series (orange), it is the system with a single definite article, for systems of the second series (yellow), the most frequent one has an indefinite article, and for the third series (green), it is the combination of a definite and an indefinite article. What these distributions show is that there are two principles that can account for the overall crosslinguistic distribution of different article systems. Firstly, systems that mark definite referents and leave the other referent types unmarked are preferred over systems that mark specific or nonspecific referents at the cost of leaving definite referents unmarked. Secondly, systems that split up the referential space according to the definite and the indefinite domain are preferred over systems that split the referential space into other domains.

With respect to article systems that mark only one referent type (in orange), this trend translates into articles marking a definite referent and leaving both specific and nonspecific referents unmarked being more frequent than articles which only mark a specific referent. Systems that would only mark a nonspecific referent and leave definite and specific referents unmarked are not attested. Table 8.41 repeats the counts of the systems in the first series together with their proportion, including the unattested system of a single nonspecific article.
Table 8.41: Observed proportions of the first system series

<table>
<thead>
<tr>
<th>system</th>
<th>N system : N systems in 1st series</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEF</td>
<td>28 : 31</td>
<td>0.90</td>
</tr>
<tr>
<td>EXSPEC</td>
<td>3 : 31</td>
<td>0.10</td>
</tr>
<tr>
<td>*NSPEC</td>
<td>0 : 31</td>
<td>0</td>
</tr>
</tbody>
</table>

Fitting the model described in (94) shows that we can be certain that the two attested systems differ in their crosslinguistic frequency. Figure 8.24 visualizes the clear difference in the estimated proportions of the two systems: languages are much more likely to have a single definite article than a single exclusive-specific article, and single nonspecific articles are extremely unlikely to be attested.

(94) A binomial regression that models the proportions of the article systems in the first series. This estimates the underlying proportions of the systems in the first series.

\[
(N \text{ system} | \text{trials}(N \text{ in 1st series}) \sim \text{system})
\]

Figure 8.24: Posterior distributions of article systems in the first series

The same trend can be observed for articles systems that mark two referent types, as was shown in Figure 8.23 (in light yellow), but in a slightly different way. The proportions for each attested system of the second series in the reduced sample are given in Table 8.42.
Table 8.42: Observed proportions of the second system series

<table>
<thead>
<tr>
<th>system</th>
<th>N system : N systems in series</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDEF</td>
<td>20 : 25</td>
<td>0.80</td>
</tr>
<tr>
<td>DEF + EXSPEC</td>
<td>5 : 25</td>
<td>0.20</td>
</tr>
<tr>
<td>INSPEC</td>
<td>0 : 25</td>
<td>0</td>
</tr>
<tr>
<td>EXSPEC + NSPEC</td>
<td>0 : 25</td>
<td>0</td>
</tr>
<tr>
<td>*DEF + NSPEC</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>*DEFNSPEC</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In this series, the most frequent system is the one that consists of a single indefinite article, marking both specific and nonspecific referents while definite referents are left unmarked. Crosslinguistically less frequent are systems with a definite and an exclusive-specific article which leave nonspecific referents unmarked and thus do not split the referential domain along the definite / indefinite domains. The same holds for systems consisting of a single inclusive-specific article, which are very rare crosslinguistically (I am not aware of languages other than Bemba with this article system). System V, with an exclusive-specific article and a nonspecific article are equally rare and not attested in the sample. This cline in frequency of systems in the second series is summarized in (95).

(95)  INDEF < DEF + EXSPEC < INSPEC < EXSPEC + NSPEC < *DEF + NSPEC, *DEFNSPEC

In (95), we see how the two principles interact and make systems with an indefinite the most frequent system in this series: if two referent types are marked, then the split into a definite domain seems to be more relevant in the majority of cases than the marking of definite referents over specific or nonspecific referents. However, if the article system does code a definite and a specific referent, leaving nonspecific referents unmarked and representing a less preferred split of the referential space, then the system with two distinct articles (DEF + EXSPEC) is more frequent than a system that uses a single article (INSPEC). This again follows the preference to single out definite referents from the other two referent types. Having two articles for specific and nonspecific referents while leaving definite referents unmarked (EXSPEC + NSPEC) makes the preferred split of the referential space into a definite and an indefinite domain. However, without the marking of definite referents by an article, it appears to be strongly dispreferred crosslinguistically to make a more fine-grained distinction in the indefinite domain and to code specific and nonspecific referents by two separate article types. In addition, the two unattested systems that would mark definite and and nonspecific referents by an article and leave specific referents unmarked violate the referent hierarchy, which can also be viewed as a strong violation of the split into a definite and an indefinite domain.
That the differences in the crosslinguistic frequency between the two most frequent systems of the second series, **INDEF** and **DEF + EXSPEC**, is statistically significant, can be confirmed by binomial regression model. The following model was fitted:

\[(N \text{ system} \mid \text{trials}(N \text{ in 2nd series}) - \text{system})\]

Figure 8.25 shows the probability distribution of the estimated proportions of the articles systems from the second series. The two non-attested systems as well as the systems **INSPEC** and **EXSPEC + NSPEC** that are attested but not part of the reduced sample are equally predicted to be very rare crosslinguisitically. More importantly, the model estimates that the system consisting of a single indefinite article is more frequent than the system of a definite with an exclusive-specific article. Therefore, we can be certain that the systems in the second series also show a clear trend towards the split into a definite and an indefinite domain.

![Figure 8.25: Posterior distributions of article systems in the second series](image)

Also for the third series of article system does Figure 8.23 (on p. 259) show that the system which splits the referential space into a definite and an indefinite domain, the system consisting of a definite and an indefinite article, is the most frequent one in the series. Table 8.43 below shows the counts together with the proportions for each system in the third series.
Table 8.43: Observed proportions of the third system series

<table>
<thead>
<tr>
<th>system</th>
<th>N system : N systems in series</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>def + indef</td>
<td>10 : 21</td>
<td>0.80</td>
</tr>
<tr>
<td>ref</td>
<td>6 : 21</td>
<td>0.20</td>
</tr>
<tr>
<td>inspec + nspec</td>
<td>3 : 21</td>
<td>0.14</td>
</tr>
<tr>
<td>def + exspec + nspec</td>
<td>2 : 21</td>
<td>0.10</td>
</tr>
<tr>
<td>*defnspec + exspec</td>
<td>0 : 21</td>
<td>0</td>
</tr>
</tbody>
</table>

The second frequent system of that series expressing all three referent types consists of a single referential article. This system does not really follow the principle that definite referents should be marked and that the referential domain should be split in a certain way; but neither does it violate those two principles. The two rarest systems in this series have an inclusive-specific with a nonspecific article and three different articles for the three different referent types. The first of those two systems, inspec + nspec violates the domain principle, as it patterns definite referents together with specific ones in contrast to nonspecific referents. The second system, def + exspec + nspec, suggests that not only those article article systems that mark definite referents in contrast to specific and nonspecific referents are crosslinguistically preferred, but that the formal distinction of specific and nonspecific referents by the use of two separate articles is actively dispreferred as well. This frequency distribution, with the most frequent type to the left, is summarized in (97). It also shows the non-attested system that belongs into this series, namely a single article to co-express definite and nonspecific referents and another exclusive-specific article. That this system is crosslinguistically unattested can again be accounted for by the general reference hierarchy that holds between the three referent types.

(97)  def + indef < ref < inspec + nspec < def + exspec + nspec < *defnspec + exspec

The following model was fitted in order to infer the underlying frequencies of the article systems in the third series:

(98)  A binomial regression that models the proportions of the article systems in the third series. This estimates the underlying proportions of the systems in the third series.
     (N system | trials(N in 3rd series) ~ system)

As we can see in Figure 8.26, the probability density functions as well as the credible intervals cover very wide areas, which is why we cannot estimate exact frequencies of the systems with a high degree of certainty. What can be seen in the relative differences is that it is very likely that the system def + indef is really is crosslinguistically more frequent than the systems def
+ EXSPEC + NSPEC and INSPEC + NSPEC (and than the unattested system as well). This is evidence supporting the overall trend that those article systems that make a split into a definite and an indefinite domain are crosslinguistically favored. Interestingly, Figure 8.26 also shows that we cannot assume the DEF + INDEF article system to be more frequent crosslinguistically than the system consisting of a single referential article.

![Figure 8.26: Posterior distributions of article systems in the third series](image)

That the two most frequent systems in this series may be equally likely to occur in a language may seem counter-intuitive, since the former is probably the most prototypical article system of all systems due to its prominence in many European languages. On the other hand, to the best of my knowledge, article systems consisting of a referential article have not been previously discussed as such. Given that we cannot be very confident of the model estimates due to the low numbers in the sample which in turn are due to the relative rarity of such complex systems in general, future work needs to show whether or not those two systems are crosslinguistically equally likely. If they are, then this shows to what extent our perception of articles is still biased by European languages.

### 8.4.2 Systems with functionally overlapping articles

What has not been addressed so far are systems with articles that overlap in their functions, e.g. a definite article together with an inclusive specific article, or an exclusive specific article with an indefinite article. In general, such systems seem rather rare crosslinguistically. In addition, there
are only two main types of systems with functionally overlapping articles found in my sample: On the one hand, we commonly find that a definite article can co-occur with a functionally more restricted anaphoric or recognitional article. This appears to be especially common in Papunesia; Table 8.44 shows the 4 languages with such an article system in the sample:

<table>
<thead>
<tr>
<th>macro area</th>
<th>language</th>
<th>article types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papunesia</td>
<td>Oksapmin</td>
<td>DEF + RECOG</td>
</tr>
<tr>
<td>Papunesia</td>
<td>Lavukaleve</td>
<td>DEF + RECOG</td>
</tr>
<tr>
<td>Papunesia</td>
<td>Urama</td>
<td>DEF + ANA</td>
</tr>
<tr>
<td>Africa</td>
<td>Jamsay</td>
<td>DEF + ANA</td>
</tr>
</tbody>
</table>

Table 8.44: Systems with overlapping articles in the definite domain

In all of these cases, the more general definite article did not fully replace the functionally more restrictive anaphoric or recognitional article. 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 (99). 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.

(99)   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)

Thus, in a system with a definite article co-occurring with an e.g. anaphoric article, the definite article can also be used in anaphoric situations, but in certain cases, e.g. the antecedent is very close or has been uttered by the same speaker, the semantically more restricted and the anaphoric article will be used, as we saw for Jamsay in section 8.3.3.2. In both Jamsay and in Urama, the two languages in the sample with this combination of articles, the anaphoric article occurs together with the definite article. Thus, we cannot really speak of “competing” markers, but of a default definite marker and the additional anaphoric article that is used to emphasize identifiability by previous mention. Examples (100) to (101) show the definite and the anaphoric articles in Urama. That =i is a definite article follows from its systematic use with anaphoric and contextually unique referents, as is shown in (100) and (101).
(100) 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 canPRS 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)

(101) [Nu moto=i] umu hiro-hia ka.
3sg house=art:def dog many-very prs
‘[In that village there lived a certain man …] At his home there were a lot of dogs.’
Urama (J. Brown et al. 2016: 85-86)

While the definite article can be used on its own to mark a referent as anaphoric, Urama makes use of aro’o in addition to the definite article to mark nouns as having anaphoric referents, as is shown in (102) below.21

(102) a. vIoro ohu=i tabo kiaukia bomo gema=i ro go’ota=i ahi ai climb top=art:def loc enough pig big=art:def nom coconut=art:def cut ka. 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 a’ai 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 of the systems that include a more general definite article and a functionally more restricted recognitional article in the sample 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 to the relevant contexts (anaphoric, contextually unique) that make it be classified as such. Therefore, it is not entirely clear whether or not the two articles functionally overlap in a strict sense or whether the “overlap” is more abstract in that the article system has one functionally broader and one more functionally more restricted article in the definite domain. Examples (103) and (104) show the use of the Oksapmin definite article jox with anaphoric and contextually unique referents, respectively. In (105), we see the recognitional article max.

21A more detailed study is needed to find the parameters on which the choice of the articles depends; it is very likely that the distance between the antecedent and the anaphoric referent plays a role though.
(103)  
a. go tap=x an pat=d=a m-p-n-gopa=li?  
\text{2sg pig=foc stay.ipfv.sg=q=emph o:prox-tell-pfv-evid.rempst.pl=report}  
Do you own a pig?  
b. \ldots jexe [tap jox] su-ti-pa=li.  
\ldots then pig art:def kill-pfv-fact.rempst.pl=report  
Then they killed the pig.

Oksapmin (Loughnane 2009: 499)

(104)  
tom jox lum p-d-m edi-pla=o [ake jox]  
water art:def a.lot caus-eat-seq stay.pfv-rempfut.sg=quot stomach art:def  
ox=o tom=wi x-ti-plox=xejox n-pli-nun.  
\text{2sg.m=quot water=only be-pfv.hodfut.sg=because o:2-tell-evid.hodpst.sg}  
\text{‘ Don’t give her too much water! Her stomach will fill up with water, she told me.” }  

Oksapmin (Loughnane 2009: 480)

(105)  
gin i ml-sa jaxe tumbuna paxna sup [\text{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)

Again, the situation in such systems may be similar to the “competition” between definite articles and demonstratives in other languages. For instance, in English, both can be used equivalently in recognitional contexts, as shown in (106b):

(106)  
a. Do you remember [the dog] (we used to have)?  
b. Do you remember [that dog] (we used to have)?  

The other type of article systems with functionally overlapping articles that are attested in the sample all involve one 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 appears important. Table 8.45 shows the languages in the sample that show such a functional overlap:

<table>
<thead>
<tr>
<th>macro area</th>
<th>language</th>
<th>article types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>Runyankore</td>
<td>inspec + ana</td>
</tr>
<tr>
<td>North America</td>
<td>Tepehua</td>
<td>inspec + indef</td>
</tr>
<tr>
<td>Eurasia</td>
<td>Basque</td>
<td>inspec + indef</td>
</tr>
<tr>
<td>Papunesia</td>
<td>Maori</td>
<td>inspec + indef</td>
</tr>
</tbody>
</table>
In all cases, it seems that the functionally broad article (inclusive-specific or referential) has extended from the definite domain to the indefinite domain, at least to specific indefinite contexts in all cases. Thus, in certain cases in which the context does not allow to disambiguate between the definite and the specific interpretation, an additional marker is used to indicate (non-)identifiability.

We will first turn to Runyankore which uses an additional marker in the definite domain in contexts in which it is important to emphasize hearer-identifiability. In Runyankore, this article is an anaphoric one, so that it only signals identifiability based on the preceding discourse. Examples (107) and (108) show the use of the inclusive-specific article in Runyankore in definite and indefinite specific contexts, respectively. To emphasize identifiability in anaphoric contexts, however, a separate anaphoric article nya is used instead in (109):

(107) **E-bi-ntu** e-bi-a o-mu-shaija a-gu
    mu-bi-t-e hangahari.
    2PL-CL8-put-IMP aside
    'Put the belongings of this man on one side.' Runyankore (Asiimwe 2014: 81)

(108) A-ha-ihi ha-ri-ho a-ba-ntu ba-ingi.
    'Nearby there are so many people.' Runyankore (Asiimwe 2014: 81)

(109) ...ahakuba nya-mu-shaija ni o-mu-ibi.
    ...because ART:ANA-CL1-man COP ART:INSPEC-CL1-thief
    ‘...for the said man is a thief.’ Runyankore (Asiimwe 2014: 86)

Tepehua, Basque, and Maori use a “mirror” strategy: instead of making definite referents more explicit in certain contexts that require disambiguation, Tepehua and Basque employ an indefinite article that partially overlaps with the inclusive-specific article, and which can disambiguate between a definite and indefinite referent by marking at as the latter.

In example (110) to (113), we see the use of the inclusive-specific articles in definite and specific indefinite contexts in Basque (110)-(111), Tepehua (112)-(113), and Maori (114)-(115), respectively.

(110) Hilabetekari horrentzat moldatu ditudan [lan purruxk-a-k]!
    monthly that.for prepared AUX.COMP work little-ART:INSPEC-PL
    'The little works that I’ve done for that monthly publication!'
    Basque (Hualde & Ortiz de Urbina 2003: 568)

(111) [Kotxe berri-a] erosi dut.
    car new-ART:INSPEC buy AUX
    'I've bought a new car.' Basque (Hualde & Ortiz de Urbina 2003: 119)
(112) [juu 7anu7 x-t’iyun-7an] lapanan waa x-ta-7asaanan. *Those two people* played instruments. Tepehua (Kung 2007: 485)

(113) t’asa-ni-kan-lich nii ka-xtaq-ni-kan-a7ch [juu lhiich’alhkat]. *They yelled that they were going to give him a job.* Tepehua (Kung 2007: 463)

(114) Ka whakamiharo anoo a Rewi ki [te kaimahi] whakahaere i te miihini uta i ngaa raakau ART:INSPEC machine load do ART:INSPEC,PL tree *Rewi marvelled again at the worker operating the machine loading the trees.* Maori (Bauer 1993: 460)

(115) Ka takai-a, ka kawe-a, ka whaka-iri-a ki runga ki [te kauere]. ART:INSPEC machine load do ART:INSPEC,PL tree *They wrapped him up and took him and suspended him in a puriri tree.* Maori (Bauer 1993: 444)

Examples (116) to (118) below show how the three languages additionally use an indefinite article in specific contexts to mark a referent as non-identifiable to the speaker (the referent of etxe ‘house’ in (116) could be either specific or non-specific; the article does not disambiguate between the two). Such a system can be motivated functionally in that it helps to reduce ambiguity in specific contexts: the inclusive specific article being vague between a definite and an indefinite specific interpretation of the referential status of the noun that it occurs with, the noun needs be interpreted as having a definite or specific referent according to other contextual cues, if available. It is exactly in those contexts in which a specific referent would rather be interpreted as a definite one, e.g. due its animacy and high-prominence argument role, that the indefinite article can be used to resolve the mismatch between the default, context-independent and the intended referential interpretation.


(117) Maa chunch nawii-ta [puma-tam kin-tata7] ... evid thus do-Pfv CL:HUMAN-ART:PRES 1pos-old.man ... ‘That’s what an old man did …’ Tepehua (Kung 2007: 631)
In addition, the articles that disambiguates between a definite and a specific indefinite interpretation of the referent in Basque, Tepehua, and Maori only partially overlap with the inclusive-specific articles in the three languages: both are indefinite articles and are used for nonspecific reference as well, as is shown in (119) and (120):

(119) Nahiz eta oso seguru-a ez den, badirudi dirulagunzeta emango digutela. 

‘Though it isn’t absolutely sure yet, it seems that they will give us a grant.’

Basque (Hualde & Ortiz de Urbina 2003: 736)

(120) [tam maqaali7] ka-maa-ch’ixtaq-ninch juu tuumiin aantu qoxiyaa 

‘A rich person could loan you money, but it isn’t good money.’

Tepehua (Kung 2007: 616)

(121) Kaaore anoo [teetahi rarururu koohahi] kia paa.

‘Not a single problem had arisen.’ 

Maori (Bauer 1993: 297)

The functional explanation of the use of an indefinite article that can disambiguate between the definite and the specific interpretation of referents in contexts in which the context-independent interpretation differs from the intended one may generally account for such a constellation of functionally overlapping articles within languages. It is important to note, however, that this does not mean that a definite article which has extended its use to specific indefinite referents necessarily triggers the development of an additional indefinite marker in a given language. Neither does it imply that the Basque, Tepehua, and Maori systems developed in this way. Especially for the two former languages, there is another evident explanation at hand: both Basque and Tepehua are in close contact with Spanish that features definite and indefinite articles. Therefore, a contact-induced development of the indefinite article in Basque and/or Tepehua should not be excluded without more detailed research in this regard. On the other hand, even if the indefinite article is a borrowing from Spanish, the fact that Basque and Tepehua had a system with a single, functional rather broad inclusive-specific article will certainly have facilitated the development of the indefinite article.
Kung (2007: 385) mentions that the use of the indefinite article in Tepehua is similar to the use of its Spanish counterpart un(os)/una(s). However, we do not seem to know more about the development of articles in Tepehua.

As for Basque, it was suggested for both the inclusive-specific article and the indefinite article that they developed in contact with Spanish. The inclusive-specific 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 inclusive-specific 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 inclusive-specific article in its early, definite stage was contact-induced, it is important that, indeed, it started as a definite article and functionally extended to mark specific indefinite referents in Modern Basque.

Concerning the development of the indefinite article in Basque, 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, its earlier form used to be 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).

Here, the relevant point concerning systems with overlapping articles is that we see that a former definite article in Basque developed into a broader inclusive-specific article in Basque. On the other hand, we see that Basque has an old plural indefinite marker that is used as an indefinite article in both the singular and the plural in Modern Basque, but that its use still appears more restricted than other neighboring indefinite articles. This means that we cannot be certain that the extension of the definite article to an inclusive-specific one caused the development of an indefinite article in Basque, but this extension, as well as a similar article in neighboring Spanish, and the availability of the partitive/indefinite plural marker batzu are three factors that certainly facilitated its development.

---

22 The inclusive-specific article -a is traditionally referred to as a definite article, as in Trask (2003: 119), de Rijk (2008: 17), and Manterola (2007). However, e.g. 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.”
8.5 Summary

This chapter gave a crosslinguistic outlook regarding the morphological properties of articles, the distribution of article types, and the types and distributions of article systems. In section 8.1, I provide selected examples to illustrate and discuss inflectional patterns of articles and show that while there are minor areal trends for the inflection of articles, we rather find a global crosslinguistic trend against the inflection of articles. For number, gender, and case marking, I showed that two factors can be singled out to restrict the inflection of the article: on the one hand, the expression of number and case on the noun appears to restrict the marking of these two categories on the article. Thus, the fact that marking gender on articles in the world’s languages is considerably more frequent than the marking of number or could be due to the general avoidance of marking the same category on more than one element in the noun phrase, manifested in the lower number of articles that mark number and case. On the other hand, I showed that articles mostly inflect for the categories that their source elements inflect for: they either retain inflection or lose the opposition of forms, but very rarely, if at all, develop new inflectional morphology.

Section 8.2 discussed two types of paradigmatic properties of articles. In the first part, I showed that we find a number of articles that formally correspond to parts of other articles within the language. These cases were shown to reflect the diachronic relation between definite and inclusive-specific articles, as well as exclusive-specific and nonspecific articles, also suggesting that nonspecific articles in systems with an inclusive-specific article developed in a different way than nonspecific articles in systems with an exclusive-specific article. The second part dealt with the definite article in Mokpe that has a tonal and a segmental exponent conditioned by the noun class. These two exponents were shown to behave differently with respect to their compatibility with possessives and demonstratives. This showed that even though the markers are exponents of a single definite article based on their referential function, there seems to be no pressure that would require different exponents of the same article to have similar morphosyntactic properties, which in turn is an argument for the independence of morphosyntactic properties of articles as referential markers.

Section 8.3 was concerned with crosslinguistic trends and variation of article types. As for areal biases, I showed that definite articles are more frequent in Africa than in other areas, while anaphoric articles are very frequent in the languages of Australia, probably due to the lack of third person pronouns that are formally distinct from demonstratives in many Australian languages. Europe and South America showed a higher number of indefinite articles than in the other areas. Regarding the overall distribution of article types, this section also showed that articles which express definite referents are significantly more frequent than articles that express either specific or non-specific referents. This finding is evidence for a split of the referential space into a definite
vs. an indefinite domain. The last part of section 8.3 discussed two article types that are hardly attested. For non-anaphoric articles, i.e. definite articles that cannot encode anaphoric referents, I proposed two scenarios that have led to their development. However, it could be shown that most of these articles are not entirely blocked in anaphoric contexts, but rather compete with another article, being used under certain circumstances. Therefore, it is questionable whether non-anaphoric definite articles exist in the strict sense. The other article type that was argued to be absent in the world’s languages was the generic article. 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 referents cannot be a primary function of these articles. In addition, due to the lack of any evidence that would point towards the existence of dedicated generic articles, I concluded that generic articles do not exist in the world’s languages.

In section 8.4, I proposed a typology of article systems on the basis of the numbers and combinations of referent types that the systems encode, i.e. how the referential space is covered by the article system. The crosslinguistic preference to encode definite articles differently from specific and nonspecific referents could also be shown to be reflected in the frequency distributions of article systems: systems that singled out definite referents compared to both specific and nonspecific referents are more frequent across languages. The distributions of article systems also pointed towards crosslinguistic gaps, all involving nonspecific articles and the hierarchy between definite, specific, and non-specific referents.

The trends discussed in sections 8.3 and 8.4 allow for five main generalizations, presented as Universals 1 to 5 below.

**Universal 1** The absence of generic articles
There are no articles whose main referential function is the coding of generic referents.

**Universal 2** The co-expression of definite and nonspecific referents
Articles can only co-express definite and nonspecific referents if they also encode specific referents.

**Universal 3** Nonspecific articles depend on specific articles
Nonspecific articles only occur in article systems with either inclusive-specific or exclusive-specific articles.

**Universal 4** The split into the definite and the indefinite domain
In the world’s languages, articles show a preference to encode definite versus specific and nonspecific referents. This manifests itself crosslinguistically in the following ways: we find a higher number of articles that encode definite referents as opposed to both specific and nonspecific referents, and we also find a preference for article systems that set apart the marking of definite
referents from the other ones, and a tendency to avoid systems that co-express definite and specific referents as well as a gap of systems that co-express definite and nonspecific referents without encoding specific referents.

**Universal 5** *Systems of articles with overlapping functions are avoided*

In the world’s languages, article systems that include articles with overlapping functions are rare. There are two types of exceptions: a language may have more than one article in the definite domain (a definite with an anaphoric or a recognitional article) or, it has a functionally broad article (inclusive-specific or referential) and an additional one that can disambiguate between the definite and the indefinite domain when important.

Universals 1 and 2 concern article types; they are expected to hold within single languages. Universal 2, excluding articles that encode definite and nonspecific referents, reflects the gaps of article system that can be accounted for by the hierarchical structure of the three main referent types (cf. Section 8.4.1). Universal 3 is implicational and restricts the existence of a nonspecific article to the existence of an inclusive-specific or exclusive-specific article in the system. It reflects the crosslinguistic gaps of systems *I* and *III*. The last universal corresponds to the crosslinguistic tendency that article systems tend to encode definite referents in contrast to both specific and nonspecific referents, which motivates the split of referential functions into a definite and an indefinite domain. This should be viewed as what is often called a “statistical” universal, i.e. a strong crosslinguistic tendency but by no means an absolute universal. The same holds for Universal 5; systems with articles that overlap in their referential functions are crosslinguistically rare, but attested in principled ways.
Chapter 9

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 a crosslinguistic discussion of the morphological properties of articles in general, the distribution of article types, and of the distribution of different types of article systems.

I showed in Chapter 2 that the first important step is a language-independent definition of articles for a crosslinguistic comparison of articles. Defining articles as adnominal referential markers, Chapter 3 proposed a detailed solution to defining referent types (referential functions) and using those as the basis for the distinction of 8 different article types that we find in the world’s languages. The article types are: definite (def), anaphoric (ana), recognitional (recog), exclusive-specific (exspec), nonspecific (nspec), indefinite (indef), inclusive-specific (inspec), and referential (ref) articles. Figure 9.1 sketches the relation between the main referent types and the 8 article types. Anaphoric and recognitional articles can be viewed as subtypes of definite articles since they encode subtypes of definite referents, namely anaphoric and recognitional ones, respectively.

![Figure 9.1: Overview of referent types and article types](image)
In addition to define different types of articles Chapter 7 addressed the issue of delimiting certain types of articles from other related elements. Especially a clear distinction between definite articles and demonstratives on the one hand and between indefinite articles and the numeral ‘one’ on the other have been problematic, since those markers are frequently related diachronically. Most approaches so far involved a list of formal and functional properties that those different elements have, describing prototypes of e.g. demonstratives vs. definite articles, which usually cannot be successfully applied to those less prototypical markers that have properties from both lists, often formal properties that would typically be related to demonstratives but a functional distribution that comes close to the one of definite articles in other languages. The approach taken in this study was to distinguish articles in the definite domain from demonstratives and articles in the indefinite domain from the numeral ‘one’ on a purely functional basis. To distinguish definite and indefinite articles from demonstratives and the numeral ‘one’, I made use of the scalar relations between different functions that both articles share with their source element. This way, cut-off points between the articles and the source elements could be defined that do not rely on form properties of the markers and that can be applied consistently across languages.

In the Chapters 4, 5, and 6, I illustrated the use of the 8 article types in detail by discussing examples from various languages, dividing the articles according to whether they belong to the definite or indefinite domain, or cross those two domains. This relates to a very important general point made in this study: definite and indefinite are not primitive referential functions. As I showed, only a more fine-grained distinction allows us to properly define and compare different articles across languages. This is not to say that the distinctions into a definite and an indefinite domain, as I propose to call them, would be unnecessary. On the contrary, while the definition of single article types relies on a more fine-grained distinction of referential functions, the crosslinguistic distributions of different article types as well as article systems can only be accounted for by this distinction.

That this distinction between a more general definite domain on the one hand and more fine-grained subtypes of definite referents is an important one also manifested itself in the properties of definite articles as shown in section 4.1: against the traditional default assumption that definite articles express all different types of definite referents, I argued that the expression of deictic referents is a consequence of the diachronic source of many definite articles. Definite articles that do not originate from demonstratives with spatial deictic functions were shown to not be used with spatial deictic referents. Instead of drawing the conclusion that such markers cannot be definite articles, the large crosslinguistic comparison of this study showed that such markers are otherwise comparable to definite articles in their use with anaphoric and contextually unique referents. Thus, instead of relying on the idiosyncratic behavior of markers with regard to deictic referents, I argued that the compatibility with deictic referents is not a criterion for definite
articles. The same was shown for absolutely unique referents, which have also often been used as an argument against classifying a definiteness marker as a definite article, in case it was not compatible with absolutely unique referents. Again, the crosslinguistic comparison of the present study showed that outside of the few well-studied definite articles in Europe, there is no trend for or against the marking of such referents by definite articles. Hence, this referent type was shown to be irrelevant as a criterion for definite articles.

Similarly, the further distinction in the indefinite domain into specific and non-specific referents allowed for a distinction of three different article types that we find in the indefinite domain: indefinite, exclusive-specific, and nonspecific articles. While it has traditionally been assumed that the development of indefinite articles includes the semantic extension from specific to non-specific referents together with the discourse-pragmatic extension from discourse-prominent to less discourse-prominent referents, I showed in section 5.3.4 that those two developments do not have to occur together. I showed for a number of indefinite articles labelled presentational articles that they can be used with both specific and nonspecific referents even though their use is restricted to discourse-prominent referents. This is certainly an issue that requires more research; within the scope of the present study, I used the notion of “discourse-prominence” 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 on the one hand, and how exactly that restricts the use of the article in single languages on the other hand. The inherent semantic properties of the referent such as animacy as well as its status as future discourse topic 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 indefinite markers within single languages.

Returning to the relevance of the distinction of a definite and an indefinite domain: although the first part of this study showed that for a proper definition and distinction of the article types we find in the world’s languages, this distinction is by no means sufficient, the second part of this study concerned with the crosslinguistic distributions of article types and article systems showed that it is nevertheless a very important one. Instead of imposing it to linguistic forms, however, this study showed that if the referential space relevant to articles is divided into the three major referent types of definite, specific, and nonspecific referents, the distinction rather emerges from the crosslinguistic distributions of both single articles and article systems. As for single articles, section 8.3.2 showed that there is a clear difference between definite referents on the one hand and specific and nonspecific ones on the other: independently from the exact article type, we found crosslinguistically significantly more articles that express definite referents than specific or nonspecific referents. Thus, this is the first piece of evidence that the split into referential domains is motivated by the crosslinguistic distribution of articles. The second piece
of evidence came from the distribution of article systems in section 8.4. While I showed that we
can distinguish a number of article systems on the basis of how they cover the referential space,
there are clear trends regarding their frequency in the world’s languages. The article systems
are repeated in Tables 9.1 to 9.3. The most frequent system of each series is highlighted in blue,
unattested systems are grayed out and marked by an asterisk. Also those preferences for and
against certain types of article systems pointed towards a preferred division into a definite and
an indefinite domain: the article systems that divide the referential space into a definite and an
indefinite domain independently from their degree of complexity were shown to be significantly
more frequent crosslinguistically than those systems that divide the referential space in another
way. The crosslinguistic gaps of article systems could be accounted for by the hierarchy between
definite, specific, and nonspecific referents in two ways. Firstly, the availability of nonspecific
articles appeared to rely on the coding of specific referents by the article system. Secondly, we
did not find any system with an article that would mark definite and nonspecific referents without
marking specific referents.

Table 9.1: First series

<table>
<thead>
<tr>
<th>referent type</th>
<th>system I</th>
<th>system II</th>
<th>system *I</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite</td>
<td>ART:DEF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>specific</td>
<td>ART:EXSPEC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nonspecific</td>
<td>ART:NSPEC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9.2: Second series

<table>
<thead>
<tr>
<th>referent type</th>
<th>system III</th>
<th>system IV</th>
<th>system V</th>
<th>system VI</th>
<th>system *II</th>
<th>system *III</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonspecific</td>
<td>ART:NSPEC</td>
<td>ART:NSPEC</td>
<td>ART:NSPEC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9.3: Third series

<table>
<thead>
<tr>
<th>referent type</th>
<th>system VII</th>
<th>system VIII</th>
<th>system XI</th>
<th>system X</th>
<th>system *IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite</td>
<td>ART:DEF</td>
<td>ART:INSPEC</td>
<td>ART:DEF</td>
<td>ART:REF</td>
<td>ART:DEFNSPEC</td>
</tr>
<tr>
<td>specific</td>
<td>ART:INDEF</td>
<td>ART:EXSPEC</td>
<td>ART:EXSPEC</td>
<td>ART:NSPEC</td>
<td>ART:EXSPEC</td>
</tr>
<tr>
<td>nonspecific</td>
<td>ART:NSPEC</td>
<td>ART:NSPEC</td>
<td>ART:NSPEC</td>
<td></td>
<td>ART:DEFNSPEC</td>
</tr>
</tbody>
</table>

The decision to include anaphoric and recognitional markers as articles was motivated by the
fact that those two types of markers, in opposition to “regular” demonstratives, are not found
in all languages and are thus similarly to definite articles non-universal, abstract grammatical
markers in the nominal domain. This is not to say that a classification of those two markers as demonstratives would not be a valid one. What this study showed by including such markers in the discussion of articles is that the availability of anaphoric and recognitional articles (or markers for that matter), clearly interacts with the availability of other types of articles. As I showed in section 8.3.1, we find an much stronger areal preference towards anaphoric articles in Australia than in all other macro areas. At the same time, other types of articles were shown to be absent in Australia, with the exception of one definite and a few recognitional articles. A less strong but potentially relevant areal trend was found for recognitional articles, which were only found in Australia and Papunesia. Whether or not one treats those markers as articles, it is clear that their presence in those two macro areas may influence the presence or rather absence of other article types. Thus, the availability of dedicated anaphoric and recognitional markers is important for a better understanding of the conditions that favor or disfavor the development of articles.

This question of how the presence / absence of one article type can influence the emergence, stability, or loss of a 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 / absence of articles; these are mainly other grammatical devices that serve the introduction and the tracking of discourse referents: verbal agreement, morphological topic and focus markers, obviative systems, complex nominal gender / class systems with gender / class marked referent tracking devices, obviative systems, etc. We know that the loss of case marking had an impact on the development of article systems in European languages. The data in this study did not provide evidence for a general crosslinguistic relation between article systems and case marking. Thus, in addition to only examining the interaction of other grammatical properties and article systems, we have to assume that those interactions are subject to areal biases. The results of this study suggest that some potential areal article-grammar interactions involve the complexity of demonstrative systems in Australia and Papunesia, differential object marking in Eurasia, and word class flexibility in Papunesia and North America, which were the two areas with the highest number of complex and typologically rare article systems.

Thus, there are still many unanswered questions especially with respect to the factors that condition the presence or absence of articles and the grammatical properties that articles interact with. This study aimed at answering some of them, laying the solid groundwork for further research on articles and article systems.
## Language Index

<table>
<thead>
<tr>
<th>Language</th>
<th>Pages</th>
<th>Language</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akan</td>
<td>98–102, 112–113</td>
<td>Ejagham</td>
<td>169</td>
</tr>
<tr>
<td>Albanian</td>
<td>34, 193–194</td>
<td>English</td>
<td>25, 37, 179, 265</td>
</tr>
<tr>
<td>Angolar</td>
<td>247</td>
<td>Fering</td>
<td>244</td>
</tr>
<tr>
<td>Arawak</td>
<td>92, 93</td>
<td>Finnish</td>
<td>27</td>
</tr>
<tr>
<td>Armenian</td>
<td>28, 239</td>
<td>German</td>
<td>19, 31, 36, 93, 172, 187, 192–193</td>
</tr>
<tr>
<td>Arrernte</td>
<td>106</td>
<td>Gooniyandi</td>
<td>105</td>
</tr>
<tr>
<td>Awakateko</td>
<td>25</td>
<td>Halkomelem</td>
<td>150–154</td>
</tr>
<tr>
<td>Ayoreo</td>
<td>254</td>
<td>Hausa</td>
<td>35</td>
</tr>
<tr>
<td>Ayutla Mixe</td>
<td>173</td>
<td>Hebrew</td>
<td>176</td>
</tr>
<tr>
<td>Bambassi</td>
<td>92</td>
<td>Hidatsa</td>
<td>123</td>
</tr>
<tr>
<td>Basque</td>
<td>239, 268</td>
<td>Indonesian</td>
<td>170, 242</td>
</tr>
<tr>
<td>Baure</td>
<td>154–162, 239</td>
<td>Irish</td>
<td>93</td>
</tr>
<tr>
<td>Bemba</td>
<td>26, 137–140, 188–189, 239, 250</td>
<td>Jamsay</td>
<td>243, 265</td>
</tr>
<tr>
<td>Biak</td>
<td>217–218</td>
<td>Kaqchikel</td>
<td>79–82, 88–90</td>
</tr>
<tr>
<td>Bilua</td>
<td>133–134</td>
<td>Kashibo-Kakataibo</td>
<td>251</td>
</tr>
<tr>
<td>Blackfoot</td>
<td>218</td>
<td>Khwarshi</td>
<td>31</td>
</tr>
<tr>
<td>Bonan</td>
<td>128–131, 240</td>
<td>Koasati</td>
<td>27</td>
</tr>
<tr>
<td>Bullom So</td>
<td>29, 239</td>
<td>Komnzo</td>
<td>102–104</td>
</tr>
<tr>
<td>Buwal</td>
<td>249</td>
<td>Konso</td>
<td>38</td>
</tr>
<tr>
<td>Carib</td>
<td>127–128</td>
<td>Lakota</td>
<td>120–121, 124, 218, 244</td>
</tr>
<tr>
<td>Ch’ol</td>
<td>122, 188, 206</td>
<td>Lango</td>
<td>93, 132–133</td>
</tr>
<tr>
<td>Chatino</td>
<td>134–136</td>
<td>Lavukaleve</td>
<td>34, 92, 265</td>
</tr>
<tr>
<td>Chinese</td>
<td>19</td>
<td>Limbum</td>
<td>94–98, 168</td>
</tr>
<tr>
<td>Coptic</td>
<td>39</td>
<td>Ma’di</td>
<td>243</td>
</tr>
<tr>
<td>Crow</td>
<td>38, 92, 123–124, 218</td>
<td>Macedonian</td>
<td>171</td>
</tr>
<tr>
<td>Dime</td>
<td>91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diyari</td>
<td>194–195</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Makah, 124
Maori, 250, 268
Mokpe, 30, 37, 82–87, 91, 247

Nivacle, 20
Nuuhchahnulth, 124
Nyaturu, 30

Oko, 240
Oksapmin, 104, 265, 266

Palula, 113, 114, 177, 240
Papuan Malay, 93
Pipil, 253
Polish, 182

Q’anjobal, 39, 109–111, 115–116, 122, 178, 218, 251

Rapa Nui, 145–149, 255
Runyankore, 250, 268

Sheko, 37, 240
Siar-Lak, 218
Slovenian, 31
Spanish, 187
Sri Lanka Malay, 29
Sumu, 28
Supyire, 189–191

Tarahumara, 93
Tepehua, 29, 142–144, 268
Tikuna, 167, 177
Tongan, 117–120, 140–142, 218
Turkish, 29
Tz’utujil, 40, 125–127

Udmurt, 19
Ughele, 26
Urama, 265
Ute, 29, 191–192

Yankuntjatjara, 105
Appendix A

Article properties (113-languages sample)
<table>
<thead>
<tr>
<th>language</th>
<th>article</th>
<th>position</th>
<th>art.infl</th>
<th>nom.cat</th>
<th>nom.infl</th>
<th>number</th>
<th>gender</th>
<th>case</th>
<th>(selected) exponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akan</td>
<td>ana</td>
<td>post</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N no</td>
</tr>
<tr>
<td></td>
<td>exspec</td>
<td>post</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N bi</td>
</tr>
<tr>
<td>Angolar</td>
<td>exspec</td>
<td>pre</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>u&quot;a N</td>
</tr>
<tr>
<td>Balanta Ganja</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N ma</td>
</tr>
<tr>
<td>Bambassi</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>if' N</td>
</tr>
<tr>
<td>Beng</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N l'ɛ</td>
</tr>
<tr>
<td>Bullom So</td>
<td>def</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N cɛ, N ācɛ, N tice, ...</td>
</tr>
<tr>
<td>Buwal</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N anta</td>
</tr>
<tr>
<td></td>
<td>exspec</td>
<td>post</td>
<td>num</td>
<td>num</td>
<td>num+gen</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N vedaye</td>
</tr>
<tr>
<td>Cape Verde Creole</td>
<td>indef</td>
<td>pre</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>un N</td>
</tr>
<tr>
<td>Dime</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num+gen+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>N-ís</td>
</tr>
<tr>
<td>Gaahmg</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>N-á</td>
</tr>
<tr>
<td>Jamsay</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N kú'á</td>
</tr>
<tr>
<td></td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>kó N</td>
</tr>
<tr>
<td>Konso</td>
<td>def</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N siʔ, N siniʔ, ...</td>
</tr>
<tr>
<td></td>
<td>exspec</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N tokka, N takka, ...</td>
</tr>
<tr>
<td>Koyra Chiini</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N di</td>
</tr>
<tr>
<td>Lango</td>
<td>pres</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N-mórō</td>
</tr>
<tr>
<td>Limbum</td>
<td>ana</td>
<td>post</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N fɔ̂</td>
</tr>
<tr>
<td>Logba</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N ɔ-kpe, ...</td>
</tr>
<tr>
<td></td>
<td>exspec</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>è N, ...</td>
</tr>
<tr>
<td>Mokpe</td>
<td>def</td>
<td>pre</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>è N, ...</td>
</tr>
<tr>
<td>Oko</td>
<td>def</td>
<td>post</td>
<td>num</td>
<td>num</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N âye, N âbe</td>
</tr>
<tr>
<td></td>
<td>exspec</td>
<td>post</td>
<td>num</td>
<td>num</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N òbèn, N ēbèn</td>
</tr>
<tr>
<td>Runyankore</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>nya-N</td>
</tr>
<tr>
<td></td>
<td>inspec</td>
<td>pre</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>a-N, o-N, ...</td>
</tr>
<tr>
<td>Sandawe</td>
<td>def</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>/</td>
<td>N-ŋ, ...</td>
</tr>
<tr>
<td>Sheko</td>
<td>def</td>
<td>post</td>
<td>gen</td>
<td>num+gen</td>
<td>/</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>N-nf, N-in</td>
</tr>
<tr>
<td>Supyire</td>
<td>def</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>/</td>
<td>N-ŋi, N-ke, ...</td>
</tr>
<tr>
<td>language</td>
<td>article</td>
<td>position</td>
<td>art.infl</td>
<td>nom.cat</td>
<td>nom.infl</td>
<td>number</td>
<td>gender</td>
<td>case</td>
<td>(selected) exponent</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
<td>----------</td>
<td>-----------------</td>
<td>---------</td>
<td>----------------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Aghul</td>
<td>indef</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>noun</td>
<td>sa N</td>
<td></td>
</tr>
<tr>
<td>Albanian</td>
<td>def</td>
<td>post</td>
<td>num+case</td>
<td>num+case</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>art</td>
<td>N-u, N-ja, ...</td>
</tr>
<tr>
<td></td>
<td>indef</td>
<td>pre</td>
<td>/</td>
<td>num+gen+case</td>
<td>case</td>
<td>/</td>
<td>art</td>
<td>noun</td>
<td>një N</td>
</tr>
<tr>
<td>Ao</td>
<td>ana</td>
<td>post</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N sə</td>
</tr>
<tr>
<td>Armenian</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>/</td>
<td>num+case</td>
<td>noun</td>
<td>noun</td>
<td>N-n</td>
<td></td>
</tr>
<tr>
<td></td>
<td>indef</td>
<td>pre</td>
<td>num+case</td>
<td>num+case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>miko N</td>
</tr>
<tr>
<td>Bantawa</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N-mko N</td>
</tr>
<tr>
<td>Barwar</td>
<td>exspec</td>
<td>pre</td>
<td>gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>xə-N</td>
<td></td>
</tr>
<tr>
<td>Basque</td>
<td>inspec</td>
<td>post</td>
<td>num+case</td>
<td>num+case</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>art</td>
<td>N-a, ...</td>
</tr>
<tr>
<td></td>
<td>pres</td>
<td>post</td>
<td>num+case</td>
<td>num+case</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>art</td>
<td>N bat, ...</td>
</tr>
<tr>
<td>Bonan</td>
<td>indef</td>
<td>post</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N-gə</td>
</tr>
<tr>
<td>Domari</td>
<td>indef</td>
<td>post</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N-ak</td>
</tr>
<tr>
<td>German</td>
<td>def</td>
<td>pre</td>
<td>num+gen+case</td>
<td>num+gen+case</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>art</td>
<td>das N, ...</td>
</tr>
<tr>
<td></td>
<td>indef</td>
<td>pre</td>
<td>gen+case</td>
<td>num+gen+case</td>
<td>num</td>
<td>noun</td>
<td>art</td>
<td>art</td>
<td>ein N, ...</td>
</tr>
<tr>
<td>Hakhun Tangsa</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>N bo</td>
</tr>
<tr>
<td>Hungarian</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>a N</td>
</tr>
<tr>
<td></td>
<td>indef</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>egy N</td>
</tr>
<tr>
<td>Irish</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>an N</td>
</tr>
<tr>
<td>Kharia</td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>mọp N</td>
</tr>
<tr>
<td>Kurtöp</td>
<td>indef</td>
<td>post</td>
<td>/</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N=the</td>
</tr>
<tr>
<td>Lepcha</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N-re</td>
</tr>
<tr>
<td>Palula</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+gen+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>so N</td>
</tr>
<tr>
<td></td>
<td>exspec</td>
<td>pre</td>
<td>/</td>
<td>num+gen+case</td>
<td>case</td>
<td>/</td>
<td>art</td>
<td>noun</td>
<td>âa N</td>
</tr>
<tr>
<td>Rajbanshi</td>
<td>exspec</td>
<td>pre</td>
<td>gen</td>
<td>num+gen+case</td>
<td>case</td>
<td>/</td>
<td>art</td>
<td>noun</td>
<td>ekta N, ...</td>
</tr>
<tr>
<td>Sri Lanka Malay</td>
<td>indef</td>
<td>flexible</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>hatthu=N</td>
</tr>
<tr>
<td>Tamil</td>
<td>indef</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>oru N</td>
</tr>
<tr>
<td>Turkish</td>
<td>indef</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>bir N</td>
</tr>
<tr>
<td>language</td>
<td>article</td>
<td>position</td>
<td>art.infl</td>
<td>nom.cat</td>
<td>nom.infl</td>
<td>number</td>
<td>gender</td>
<td>case</td>
<td>(selected) exponent</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>-----------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Abui</td>
<td>recog</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N nu</td>
</tr>
<tr>
<td>Agta</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>i N</td>
</tr>
<tr>
<td></td>
<td>exspec</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N=a</td>
</tr>
<tr>
<td>Bajau</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N e</td>
</tr>
<tr>
<td>Biak</td>
<td>exspec</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>/</td>
<td>N=ya, N=na, ...</td>
</tr>
<tr>
<td></td>
<td>def</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>/</td>
<td>N anya, N ani, ...</td>
</tr>
<tr>
<td></td>
<td>nspec</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>N=o, N=no</td>
</tr>
<tr>
<td>Bilua</td>
<td>pres</td>
<td>pre</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>/</td>
<td>kala N</td>
</tr>
<tr>
<td>Bunaq</td>
<td>ana</td>
<td>post</td>
<td>gen</td>
<td>gen</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>N ba</td>
</tr>
<tr>
<td>Komnzo</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+gen+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>ane N</td>
</tr>
<tr>
<td></td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>num+gen+case</td>
<td>num+gen+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>nā N</td>
</tr>
<tr>
<td>Lavukaleve</td>
<td>def</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num+gen</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N na, N la, ...</td>
</tr>
<tr>
<td></td>
<td>indef</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num+gen</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N roa, N ro, ...</td>
</tr>
<tr>
<td></td>
<td>recog</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num+gen</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N me, N mea, ...</td>
</tr>
<tr>
<td>Maori</td>
<td>inspec</td>
<td>pre</td>
<td>num</td>
<td>num+case</td>
<td>case</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>noun</td>
</tr>
<tr>
<td>Menya</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N=i</td>
</tr>
<tr>
<td>Oksapmin</td>
<td>indef</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N tit</td>
</tr>
<tr>
<td></td>
<td>recog</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N max</td>
</tr>
<tr>
<td></td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N jox</td>
</tr>
<tr>
<td>Papuan Malay</td>
<td>def</td>
<td>post</td>
<td>num</td>
<td>num</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>N diu</td>
</tr>
<tr>
<td>Rapa Nui</td>
<td>ref</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>te N</td>
</tr>
<tr>
<td>Siar Lak</td>
<td>inspec</td>
<td>pre</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>/</td>
<td>ep N, a N, ...</td>
</tr>
<tr>
<td></td>
<td>nspec</td>
<td>pre</td>
<td>gen</td>
<td>num+gen</td>
<td>/</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>ti N, tok N, ...</td>
</tr>
<tr>
<td>Sunwadia</td>
<td>ref</td>
<td>pre</td>
<td>num+case</td>
<td>num+case</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>a=N, na=N, ...</td>
</tr>
<tr>
<td>Sye</td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>hai N</td>
</tr>
<tr>
<td>Teiwa</td>
<td>ana</td>
<td>post</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N waal</td>
</tr>
<tr>
<td></td>
<td>pres</td>
<td>post</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N nuk</td>
</tr>
<tr>
<td>Tongan</td>
<td>inspec</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>he N</td>
</tr>
<tr>
<td></td>
<td>nspec</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>ha N</td>
</tr>
<tr>
<td>Urama</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N-i</td>
</tr>
<tr>
<td></td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>noun</td>
</tr>
</tbody>
</table>

Table A.3: Article properties (Papunesia)
<table>
<thead>
<tr>
<th>language</th>
<th>article</th>
<th>position</th>
<th>art.infl</th>
<th>nom.cat</th>
<th>nom.infl</th>
<th>number</th>
<th>gender</th>
<th>case</th>
<th>(selected) exponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrernte</td>
<td>ana</td>
<td>post</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>nhenge N</td>
</tr>
<tr>
<td>Bardi</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>noun</td>
<td>/</td>
<td>jarri N</td>
</tr>
<tr>
<td>Bininj Kun-Wok</td>
<td>recog</td>
<td>post</td>
<td>gen</td>
<td>num+gen</td>
<td>/</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>nawu, ngalu N</td>
</tr>
<tr>
<td>Diyari</td>
<td>def</td>
<td>pre</td>
<td>num+gen+case</td>
<td>num+gen+case</td>
<td>case</td>
<td>art</td>
<td>art+noun</td>
<td>nhan N, ...</td>
<td></td>
</tr>
<tr>
<td>Gooniyandi</td>
<td>recog</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>noun</td>
<td>ginharndi N</td>
</tr>
<tr>
<td>Guragone</td>
<td>ana</td>
<td>pre</td>
<td>gen</td>
<td>gen</td>
<td>/</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>adjini N, ...</td>
</tr>
<tr>
<td>Jingulu</td>
<td>ana</td>
<td>pre</td>
<td>num+gen+case</td>
<td>num+gen+case</td>
<td>case</td>
<td>art</td>
<td>art+noun</td>
<td>kun N, ...</td>
<td></td>
</tr>
<tr>
<td>Mangarrayi</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+gen+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>noun</td>
<td>ginara N</td>
</tr>
<tr>
<td>Martuthunira</td>
<td>ana</td>
<td>pre</td>
<td>case</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>art+noun</td>
<td>/</td>
<td>nhiyaa N</td>
</tr>
<tr>
<td>Wardaman</td>
<td>ana</td>
<td>post</td>
<td>/</td>
<td>num+gen+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>N-marlang</td>
</tr>
<tr>
<td>Warrwa</td>
<td>ana</td>
<td>pre</td>
<td>case</td>
<td>num+case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>art</td>
<td>kinya N</td>
</tr>
<tr>
<td>Wubuy</td>
<td>ana</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>art+noun</td>
<td>/</td>
<td>/</td>
<td>N bagi, ...</td>
</tr>
<tr>
<td>Yankuntjatjara</td>
<td>recog</td>
<td>post</td>
<td>/</td>
<td>num+case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>noun</td>
<td>panya N</td>
</tr>
<tr>
<td>Yolŋu</td>
<td>ana</td>
<td>pre</td>
<td>num</td>
<td>num+gen+case</td>
<td>num+case</td>
<td>art+noun</td>
<td>/</td>
<td>noun</td>
<td>nunhi N, ...</td>
</tr>
<tr>
<td>Yuwaalaraay</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>ngaama N</td>
</tr>
<tr>
<td>language</td>
<td>article</td>
<td>position</td>
<td>art.infl</td>
<td>nom.cat</td>
<td>nom.infl</td>
<td>number</td>
<td>gender</td>
<td>case</td>
<td>(selected) exponent</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Chatino</td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ N</td>
</tr>
<tr>
<td>Chimariko</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N-&lt;i&gt;op&lt;/i&gt;</td>
</tr>
<tr>
<td>Chocholtec</td>
<td>pres</td>
<td>pre</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ N</td>
</tr>
<tr>
<td>Ch'ol</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/ li N</td>
</tr>
<tr>
<td>Chumash</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ lo&lt;sup&gt;‘&lt;/sup&gt;ka-N</td>
</tr>
<tr>
<td>Chumash</td>
<td>indef</td>
<td>pre</td>
<td>gen</td>
<td>/</td>
<td>num+gen</td>
<td>/</td>
<td>/</td>
<td>art</td>
<td>/ jump&lt;sup&gt;‘&lt;/sup&gt;ej N, ...</td>
</tr>
<tr>
<td>Cupeño</td>
<td>def</td>
<td>pre</td>
<td>num</td>
<td>num+case</td>
<td>num+case</td>
<td>art+noun</td>
<td>noun</td>
<td>/</td>
<td>/ pe N, pem N</td>
</tr>
<tr>
<td>Haida</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ N-aay</td>
</tr>
<tr>
<td>Halkomelem</td>
<td>ref</td>
<td>pre</td>
<td>gen</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>art</td>
<td>/</td>
<td>/ kwʔa N, tə N</td>
</tr>
<tr>
<td>Koasati</td>
<td>ana</td>
<td>post</td>
<td>/</td>
<td>num+case</td>
<td>num+case</td>
<td>noun</td>
<td>/</td>
<td>noun</td>
<td>N-&lt;i&gt;sáya&lt;/i&gt;</td>
</tr>
<tr>
<td>Lakota</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ N ki</td>
</tr>
<tr>
<td></td>
<td>exspec</td>
<td>post</td>
<td>num</td>
<td>num</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>N wâ</td>
</tr>
<tr>
<td></td>
<td>nspec</td>
<td>post</td>
<td>num</td>
<td>num</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>N wâži</td>
</tr>
<tr>
<td>Nuuchahnulth</td>
<td>inspec</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>/ N-ʔi</td>
</tr>
<tr>
<td>Pipil</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>/ ne N</td>
</tr>
<tr>
<td></td>
<td>indef</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>/ se: N</td>
</tr>
<tr>
<td>Sahaptin</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+case</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ kw’ink N</td>
</tr>
<tr>
<td>Tarahumara</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ echi N</td>
</tr>
<tr>
<td>Tepehua</td>
<td>inspec</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ juu N</td>
</tr>
<tr>
<td></td>
<td>indef</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ tam N</td>
</tr>
<tr>
<td>Ulwa</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>/ N ya</td>
</tr>
<tr>
<td></td>
<td>pres</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>/ N as</td>
</tr>
<tr>
<td>Ute</td>
<td>def</td>
<td>post</td>
<td>num+gen+case</td>
<td>num+gen+case</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>art</td>
<td>N ’uru, ...</td>
</tr>
<tr>
<td>Yurok</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/ kue N</td>
</tr>
<tr>
<td>Zoque</td>
<td>def</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>/ bi N</td>
</tr>
<tr>
<td>language</td>
<td>article</td>
<td>position</td>
<td>art.infl</td>
<td>nom.cat</td>
<td>nom.infl</td>
<td>number</td>
<td>gender</td>
<td>case</td>
<td>(selected) exponent</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>---------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Aguaruna</td>
<td>ana</td>
<td>pre</td>
<td>case</td>
<td>case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>art+noun</td>
<td>nu N</td>
</tr>
<tr>
<td>Akuntsú</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N-(e)t</td>
<td></td>
</tr>
<tr>
<td>Apinayé</td>
<td>def</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>N ja</td>
</tr>
<tr>
<td>Arawak</td>
<td>def</td>
<td>pre</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>li N, ...</td>
</tr>
<tr>
<td></td>
<td>indef</td>
<td>pre</td>
<td>/</td>
<td>num+gen</td>
<td>/</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>aba N</td>
</tr>
<tr>
<td>Asheninka Perené</td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>aparoni N</td>
</tr>
<tr>
<td>Ayoreo</td>
<td>inspec</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>nonspec</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>art</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Baré</td>
<td>ana</td>
<td>pre</td>
<td>gen</td>
<td>num+gen</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>da N</td>
</tr>
<tr>
<td>Baure</td>
<td>ref</td>
<td>pre</td>
<td>gen</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>art</td>
<td>/</td>
<td>to N, ti N</td>
</tr>
<tr>
<td>Carib</td>
<td>indef</td>
<td>pre</td>
<td>num</td>
<td>num</td>
<td>/</td>
<td>art</td>
<td>/</td>
<td>/</td>
<td>N amu</td>
</tr>
<tr>
<td>Cubeo</td>
<td>ana</td>
<td>pre</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>di=N</td>
</tr>
<tr>
<td>Kashibo-Kakataibo</td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>case</td>
<td>case</td>
<td>/</td>
<td>/</td>
<td>noun</td>
<td>achusi N</td>
</tr>
<tr>
<td>Mamaindé</td>
<td>ref</td>
<td>post</td>
<td>/</td>
<td>num</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N-ã, N-ču</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>kiñe N</td>
</tr>
<tr>
<td>Nhengatu</td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>yepe N</td>
</tr>
<tr>
<td>Parecis</td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>num</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>hatya N</td>
</tr>
<tr>
<td>Sabanê</td>
<td>ref</td>
<td>post</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>N-mi</td>
</tr>
<tr>
<td>Wayuu</td>
<td>def</td>
<td>post</td>
<td>num+gen</td>
<td>num+gen</td>
<td>num</td>
<td>art+noun</td>
<td>art</td>
<td>/</td>
<td>N-kalü, N-kai, ...</td>
</tr>
<tr>
<td></td>
<td>pres</td>
<td>pre</td>
<td>/</td>
<td>num+gen</td>
<td>num</td>
<td>noun</td>
<td>/</td>
<td>/</td>
<td>wane N</td>
</tr>
<tr>
<td>language</td>
<td>type</td>
<td>inflection</td>
<td>behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akan</td>
<td>ana</td>
<td>n</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buwal</td>
<td>def</td>
<td>n</td>
<td>different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jamsay</td>
<td>def</td>
<td>n</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Konso</td>
<td>def</td>
<td>num</td>
<td>different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logba</td>
<td>def</td>
<td>n</td>
<td>different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oko</td>
<td>def</td>
<td>num</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runyankore</td>
<td>ana</td>
<td>n</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albanian</td>
<td>def</td>
<td>num+case</td>
<td>different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armenian</td>
<td>def</td>
<td>n</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basque</td>
<td>inspec</td>
<td>num+case</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>def</td>
<td>num+gen+case</td>
<td>different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td>def</td>
<td>n</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palula</td>
<td>ana</td>
<td>n</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agta</td>
<td>def</td>
<td>n</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biak</td>
<td>exspec</td>
<td>n</td>
<td>different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Komnzo</td>
<td>ana</td>
<td>n</td>
<td>same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>language</th>
<th>type</th>
<th>inflection</th>
<th>behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lavukaleve</td>
<td>def</td>
<td>num+gen</td>
<td>same</td>
</tr>
<tr>
<td>Oksapmin</td>
<td>recog</td>
<td>n</td>
<td>same</td>
</tr>
<tr>
<td>Siar Lak</td>
<td>inspec</td>
<td>num+gen</td>
<td>different</td>
</tr>
<tr>
<td>Teiwa</td>
<td>ana</td>
<td>n</td>
<td>same</td>
</tr>
<tr>
<td>Tongan</td>
<td>inspec</td>
<td>n</td>
<td>same</td>
</tr>
<tr>
<td>Urama</td>
<td>def</td>
<td>n</td>
<td>same</td>
</tr>
<tr>
<td>Chol</td>
<td>def</td>
<td>n</td>
<td>different</td>
</tr>
<tr>
<td>Chumash</td>
<td>def</td>
<td>n</td>
<td>same</td>
</tr>
<tr>
<td>Lakota</td>
<td>def</td>
<td>n</td>
<td>different</td>
</tr>
<tr>
<td>Pipil</td>
<td>def</td>
<td>n</td>
<td>same</td>
</tr>
<tr>
<td>Tepehua</td>
<td>inspec</td>
<td>n</td>
<td>same</td>
</tr>
<tr>
<td>Ulwa</td>
<td>def</td>
<td>n</td>
<td>same</td>
</tr>
<tr>
<td>Arawak</td>
<td>def</td>
<td>num+gen</td>
<td>different</td>
</tr>
<tr>
<td>Ayoreo</td>
<td>inspec</td>
<td>num+gen</td>
<td>same</td>
</tr>
<tr>
<td>Wayuu</td>
<td>def</td>
<td>num+gen</td>
<td>different</td>
</tr>
</tbody>
</table>

Table A.7: Inflectional behaviour of different articles within single languages
Appendix B

Examples

This appendix contains the relevant examples of each article in the sample. For anaphoric and recognitional articles, those correspond to anaphoric and recognitional contexts, respectively. Definite articles are shown with anaphoric as well as contextually unique referents. For exclusive-specific articles and nonspecific articles, I provide examples with specific and nonspecific referents each; indefinite (and presentational) articles are shown with both referent types. The two domain-crossing article types are illustrated accordingly: inclusive-specific articles are shown in three examples with an anaphoric, contextually-unique, and specific referent, and for referential articles, a context with a nonspecific referent is given in addition to those other three contexts.

B.1 Africa

Akan

Anaphoric article

(1) A: me-kɔ-duru-uɛ no, me-hu-u ɔbaa bi ne mmɔfra mmienu
1SG-go-arrive-PST CD 1SG-see-PST woman ART:EXSPEC with children two
bi se wo-te mukaasese hɔ.
ART:EXSPEC COMP 3PL-sit kitchen there
‘When I arrived, I saw a woman with two children sitting in the kitchen.’

B: (Aso) wo-nim [ɔbaa no] (anaa)?
Q 2SG-know woman ART:ANA Q
‘Did you know the woman?’ (prim. data)
Exclusive-specific article

Specific referent:

(2) Amma ka-a [akura bi] hu asën kyere-ɛ me. ye-n-kɔ hwe ho Ama tell-pst village ART:EXSPEC self matter show-pst 1sg 1pl-imp-go see there ‘Ama told me about a (certain) village. Let’s go and see.’ (prim. data)

Absence with a nonspecific referent:

(3) me-re-pɛ [ɔbɔmmɔfoɔ *bi] aa ɔ-bɔ adwuma te se me 1sg-prog-like hunter ART:EXSPEC REL 3sg-hit work be like 1sg ‘I am looking for a (any) hunter who works as hard as I do.’ (prim. data)

Angolar (Exclusive-specific article)

Specific referent:

(4) N ka tango [u*a ɔoya]. 1sg fut tell ART:EXSPEC story ‘I will tell a story.’ (Lorenzino 1998: 133)

Absence with a nonspecific referent:

(5) Ši n ka vuna kikie 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.’ (Lorenzino 1998: 170)

Balanta Ganja (Definite article)

Anaphoric referent:

(6) a. bi-ɲaŋ batʃi fembɛ nduba tu reŋu. cl2-people village dem:cl5 all all meet ‘The people of that village got together.’

b. bi-gi niŋi niRE-e-m bala NDANŋ. s:3pl-cop with dance-gen balafon big ‘They were having a big balafon dance.’

c. [bi-ɲaŋ ma] jalu niRE-e. cl2-people ART:DEF begin dance ‘The people began to dance.’ (Fudeman 1999: 351)
Contextually unique referent:

(7) ma bi-santu ngi [bi-kalfa ma].
cons s:3PL-talk with cl2-important.person ART:DEF
‘And then they talked with the most important men of the village.’ (Fudeman 1999: 356)

Bambassi (Definite article)

Anaphoric referent:

(8) a. jenif ḍur-èt hiḵk-ògor-iʃ bɪʃ-òw-á.
ancient year-LOC one-chief-sbj exist-pst:hab-decl
‘In the old days, there used to be a chief.’

b. [iʃ ògor] mûnts’-iʃ nû: wiː-gà-m-bîʃ-à:
ART:DEF chief woman-sbj how say-fut-3-npst:aux-intr
‘And what does a wife of the chief say?’
(Ahland 2012: 284)

Contextually unique referent:

eat:imper:2pl say-ss:nb eat-thing-obj 3-pl-obj 1sg-give-decl
‘I said “Eat!”, and gave them the food.’

b. tí-ʃ [iʃ kâl-lâ] ha-tí-mī-1â wiː-in i-té iʃ-kol-lâ
1sg-sbj ART:DEF porridge-obj aff-1sg eat-decl say-ss:nb 3sg-sbj 3-pl-obj
ha-mé:nt-1â
aff-tell-decl
‘She told them, saying “I ate the porridge”.’
(Ahland 2012: 640)

Beng (Definite article)

Anaphoric referent:

(10) Zrē fɛ mî-ò à yé mi mî [zrē bi-ɛ] yà.
road rel 2sg-stat 3sg mouth 2sg:hab 2sg road this ART:DEF walk
‘The road you stand on, walk this road.’ (Paperno 2014: 102)

Contextually unique referent:

(11) ganmlàn, [min suø lè] o lè geŋ.
chimpanzee 2sg house ART:DEF 3sg.pst cop beautiful
‘Chimpanzee, your house is nice.’ (Paperno 2014: 109)
Bullom So (Definite article)

Anaphoric referent:

(12) là ɲà ké sú sì-čé ḥá kámpɔ̀r [sú sì-čé] kò kùŋk-ɛ̀
when 3pl see fish CL-ART:DEF 3pl enclose fish CL-ART:DEF to fence-in
‘When they see the fish, they enclose the fish with the net.’ (Childs 2011: 242)

Contextually unique referent:

(13) a. tàmrɔ̀ì ŋ-wɔ́ kò títikɛ̀i.
    tamro EMPH-3sg go 3sg descend to Titike
    ‘Tamro left, he went down to Titike.’

b. [ũ-sàl cɛ́] kò cè nún ã-bömùn bɔ̀lù ù tòl
    CL-river ART:DEF it COP now CL-big and 3sg descend
    ‘The river was still high so he went down (somewhere else).’
    (Childs 2011: 245)

Buwal

Definite article

Anaphoric referent:

(14) Hwa-baw taf mɑ=kɛ̀da uza, [taf anta] a-baw, a-nda ta ḥ
    s:2sg-turn path REL=towards down [path ART:DEF] s:3sg-turn s:3sg-go through PREP
    school
    ‘You turn on the path which goes down, the path turns, it goes through the school.’
    (Viljoen 2013: 239)

Contextually unique referent:

(15) Ca hwa-nda, hwa-ja aza [kɑŋɡaŋ ñkwa anta] ara ca vɑŋgɛ̀?
    TOP 2SG.SBJ-come 2SG.SBJ-hit SRC drum POSS:2SG ART:DEF SIM TOP how
    ‘How can you come here while playing your drum?’
    (Viljoen 2013: 240)
Exclusive-specific article

Specific referent:

(16) [ŋhwə-ye vedaye] a wata juraw ete a Zukwadfanw aka.
    goat-pl ART:EXSPEC.PL PREP compound sub.chief here PREP Zukodfong EXIST
    ‘There are some goats at the sub-chief’s home here in Zukodfong.’ (Viljoen 2013: 455)

Absence with a nonspecific referent:

(17) Berjeŋ naka akwaw.
    donkey POSS:1SG NEG.EXIST
    ‘I don’t have a donkey.’ (Viljoen 2013: 490)

Cape Verde Creole (Indefinite article)

Specific referent:

(18) N ten [un duensa] ki’ N dja ara ku el, ta txoma asma duensa
    I have ART:INDEF disease COMP I PFV bother with it NFUT call asthma disease
    ta txoma asma i mi, nha trabadju, azagua, N ta trabadja mas es duensa
    NFUT call asthma and me my work rainy.season I NFUT work but DEM disease
    p’e trabadja-m sabi e ku txuba.
    for.it work-me well COP with rain
    ‘I have a disease that bothers me, they call it asthma, the disease is called asthma and I do
    my work harvesting but for this disease not to bother me, it must rain’
    (Baptista 2002: 31-32)

Nonspecific referent:

(19) Nu ta ranja [un panela], nu ta ba kunznha la n’otu funkü.
    we IRR get ART:INDEF pan we IRR go cook there in other hut
    ‘We find a way of getting a pan, then we go and cook in the hut over there.’
    (Baptista 2002: 26)
Dime (Definite article)

Anaphoric referent:

(20) a. kên-ká yer-ká wôkkil ?insé wôtú kub-ó gaš-ká tîŋ-á bay-im
dog-conj donkey-conj one day s:1pl forest-loc road-instr go-cv food-acc
k’áy-á ?íst-tûb ?énet-’á ...
find-cv eat-fut say-cv ...
‘Once upon a time, a dog and a donkey agreed and say ...’

b. k’ôt-úb bow-de yer-ís šün-im ŋits-á gi-gís’-i-t.
arrive-rel.m dir-abl donkey-art:def grass-acc eat-cv red-satisfy-pfv-1
‘The donkey found grass to eat and satisfied himself.’ (Seyoum 2008: 180-181)

Contextually unique referent:

(21) šiftay-ko šif-is ?een nît-s-ôb
shiftaye-gen shoes-art:def early childhood
‘The shoes of Shiftaye are from his childhood.’ (Seyoum 2008: 45)

Gaahmg (Definite article)

Anaphoric referent:

(22) a. Bâárg=á âò-âm án-ân ì jëgg Gôôr=ë.
Baggara=art:def coming staying with people Goor=with
‘The Baggara were coming with the people of Goor.’

b. …[jëgg Gôôr=ë] ì bá ás-ì ággân jëgg ìn-g=ï
…people Goor=art:def oh became=for.us things bad-pl=rc
‘...The Goor tribe became our enemies [lit. bad things to us].’ (Stirtz 2011: 340)

Contextually unique referent:

(23) ñëñ ì billì ì de kùnḑ, mô=x ì bâd-d gôôl=î
person s:3sg shot:6:3sg poss:3sg chest fire=art:def s:3sg penetrate-inf shield=art:def
lâŋ tú ì ñîr-ë.
until out s:3sg die-inf
‘A person shot him in his chest through the shield so that he died.’ (Stirtz 2011: 330)
Jamsay

Definite article

Anaphoric referent:

(24) nǐː tūrú kɛŋ nɔː, kɔŋ nɔː mɛy nɔː kɛŋ [kɔŋ day one t̪ɔp mɛlɛ办好 drink.PVF.3SG mɛlɛ办好 drink and until go and mɛlɛ办好 kʊ̞] kɔ̀ŋ àː

ART·DEF NONH.O catch.PVF.3SG

‘One day, he [the Mouse] drank some millet beer, he drank the millet beer, to the point that the millet beer grabbed him [made him drunk].’ (Heath 2008: 256)

Contextually unique referent:

(25) ū-jũwɔ kɔ̞ nāː kʊ̞ jẽ́ mɛy [gũ̞ n kʊ̞] diɗ̄ nũ-yɛ̀

RED-MOUSE DEM food ART·DEF for and bowl.cover ART·DEF follow enter-PVF.S:3SG

tāŋā dẽ̀,...
happen if

‘The mouse, when it has followed along the bowl cover and has gone in (under the bowl itself) to get the food, …’ (Heath 2008: 302)

Anaphoric article

(26) tɔ̞gū pɔːrɔ tɔ̞gɔ̀ kɔ̀ bẽ̀rɛ́ bẽ̀ nũː [kɔ̀ tɔ̞g kʊ̞] ùrɔ̀

shed first shed.building 3SG.NHUM in 3PL enter.PVF ART·ANA shed ART·DEF house

tān̄-ŋā mɛy̞ nĩŋ yɔ̞=kɔ̀

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.’ (Heath 2008: 164)

Konso

Definite article

Anaphoric referent:

(27) a. luq̄is̄a kι̞̂-eeta uwwaa parraa a ʃatt̄-a̞ye akata kokkook-aa?.

leather.skirt say-INF skirt years GEN long.time.ago-top very strong-pl

‘Many years ago leather skirt meant an expensive skirt.’
b. ee luqūsa-siʔ piddf-t-i oore-neeyye maana=i kod-t-i tuparraa
when leather.skirt-ART:DEF buy-3.f-PFV then-TOP what=3 do-3.f-PFV girls
akki=i pīfaaʔ erk-t-i.
two=3 water-DAT send-3.f-PFV
‘When she bought the leather skirt, what did she do? She sent two girls to fetch water.’
(Orkaydo 2013: 281-282)

Contextually unique referent:

(28) xala gīmaya-sik karmaa i-iff-ay
yesterday old.man-ART:DEF lion 3-kill-PFV.3.M
‘Yesterday the old man killed a lion.’ Konso (Orkaydo 2013: 198)

Exclusive-specific article

Specific referent:

(29) ifeeddai haad-t-i haad-t-i inanta-siʔ ka irroota sakal tuull-f-t-i
then carry-3.f-PFV carry-3.f-PFV girl-ART:DEF and mountain nine cross-CAS-3.f-PFV
ka a kudanttat-eeyye [jila tokka] kapa-a kay-i-n karmaa-siʔ
and GEN ten-TOP rock ART:PRES.M near-LOC reach-PFV-PL lion-ART:DEF
i=kid-ni: ayikka tika-awo-y ana lekkīf-i.
3=say-IPFV-PRES here house-POSS:1SG-TOP 1SG.ACC step.down-PFV
‘Then, the girl carried and carried the lion over nine mountains and on the tenth one near
a rock, the lion said: Let me down as my house is here.’ (Orkaydo 2013: 289)

Absence with a nonspecific referent:

(30) alleeta-siʔ a=i tuuda kela=in kit-ni-n malla=i piʔ-t-i.
hut-ART:DEF that=3 pillar under=NEG.3 be-PRES.PFV-NEG because=3 fall-3f-PFV
‘The hut collapsed because there is no pillar under it.’ (Orkaydo 2013: 240)

Koyra Chiini (Definite article)

Anaphoric referent:

(31) a. i hisa ka din gandoo alkaasu, i faraa-ndi gi nda laamu.
s:3PL do.much INF take this.land tax s:3PL suffer-CAS O:3PL with rule
‘They [the Tuaregs] took a great deal of this land’s taxes, they oppressed them [local
people] with their iron rule.’
b. i din [alkasu di] hal i hisa ka faraa-ndi boro di yo.
s:3PL take tax ART:DEF until s:3PL do.much INF suffer-CAUS person ART:DEF PL
‘They took taxes to the point that they oppressed the people very much.’

(Heath 1999: 434)

Contextually unique referent:

(32) maabe di jow ñgu wane [yenje jiney di] yo.
griot ART:DEF take refl:3sg poss fight implement ART:DEF PL
‘The griot took his battle gear.’

(Heath 1999: 439)

Lango (Presentational article)

Specific referent:

(33) pünk-ɔ́rɔ̀ tië i ñè ɔ̀t
pig-ART:PRES be.present.HAB.3SG back ɔ́t house
‘There is a pig behind the house.’

(Noonan 1992: 162)

Nonspecific referent:

(34) cën-nɔ́rɔ̀ pé
ghost-ART:PRES be.absent.HAB.3SG
‘There are no ghosts.’

(Noonan 1992: 147)

Limbum (Anaphoric article)

(35) a. ñwè mɔ̀ à m có m ɔ̀t
man one 3SG PST3 PROG live in Nkambe
‘There used to live a man in Nkambe.’

b. [ñwè ṭɔ̀]
man ART:ANA 3SG PST3 have children seven
‘The man had seven children.’

(prim. data)

Logba

Definite article

Anaphoric referent:

(36) a. ekple abó-yayi ǝdjì.
then 2SG-FUT-search soda
‘Then you will search for soda.’
b.  adj=é ɔ-mɛ oɖu ivi-kpi=é xé atsi-tsi-ŋú kókó afɔtsi soda=ART:DEF AM-DEM 3SG-be thing-one=ART:DEF REL 1PL-HAB-see coca pod tsúeyi aló bladzo afɔtsi kpɛ ibikpa zuzz-go nu. dry or plantain peel CONJ palm.husk roast-nmlz in ‘We get the soda from cocoa pod which is dry or plantain peel and palm husk which is roasted.’ (Dorvlo 2008: 345-346)

Contextually unique referent:

(37) i-ta-tɛ́ atsú etsi=é xé i-ɖu Ghana umɛ koko 3SG-give-compl 1PL land=ART:DEF REL 3SG-be Ghana here cocoa mo-ό-nyɛ-n-zí ɔdzɔgbe nu. NEG-3SG-stay-NEG-well savanna in ‘It is that in our land here, in Ghana, cocoa does not do well in the savanna.’ (Dorvlo 2008: 350)

Exclusive-specific article

Specific referent:

(38) e-bìtsi-ɛ́ ɔ-kla fɛ́ [ɔ-yɔ́ cl7-tree mango ɔ-kpie] dzáa ko etsi cl5-child-ART:DEF s:SG-hide into cl7-tree big cl7-ART:PRES no.addition only under ‘The child went and hid under only a big tree.’ (Dorvlo 2008: 93)

Absent with a nonspecific referent:

(39) xé a-bɔ́-za nfú, gbã ibotɛ́ te á-yayi asɔ alo gaze. if 2SG-FUT-cook oil first reason compl 2SG.FUT-search pot or iron.pot ‘If you want to make palm oil, first you have to search for a pot or an iron pot.’ (Dorvlo 2008: 331)

Mokpe (Definite article)

Anaphoric referent:

(40) B: málúwá má βél-ő kíʒení water.cl6.art:def cl6 is-in kitchen ‘There is water in the kitchen.’
A: ðɔá ńɛ́ nà ʒéni málúwá/ *màluwa come please I see.NEG water.cl6.art:def water.cl6 ‘Please come, I cannot find the water.’ (prim. data)
Contextually unique referent:

(41) [è mòkànèlì] à 3òb-è bóndá jà è ndʒùmá
\text{ART:DEF} leader.CL1 CL1 NEG-have time for \text{ART:DEF} fight.CL9

‘The president does not care about the strike.’

literally: ‘The leader does not have time for the fight.’ \hfill (prim. data)

\textbf{Oko}

\textbf{Definite article}

Anaphoric referent:

(42) a. Àkọ bi-sie a-gá-è na, àyè bẹ-gba ka óró óbèn a-ca
as s:3PL-do PROG-say-it RC FOC s:3PL-see COMP person ART:INDEF.SG PROG-come
nènè file àlègbẹ nènè rùm nà.
REL wear cloth REL thick RC

‘As they were saying it, so they saw someone coming who was wearing a thick cloak.’

b. àyè bẹ-na ka óró nènè i-bi-ìwú nènè àkà-tàyè dì
FOC s:3PL-accept COMP person REL LOC-poss:3PL-body REL FUT-do.before can
wọrè ka óró àyè e-tie [àlègbẹ àyè]
caus REL COMP person ART:DEF.SG s:3SG-take cloth ART:DEF.SG do.away RC COMP 3SG
e-sű i-ìwú-ógbìgbìn fọrẹ nè.
REL have strength surpass for

‘So they agreed that anyone among them, who was able to make the man pull off the
cloak, was stronger.’ \hfill (Atoyebi 2010: 274-275)

Contextually unique referent:

(43) ìkèn óbèn wàmò nènè è-gbè ka [údúdó àyè] á-wà èrùnrò
town ART:EXSPEC.SG exist REL s:3SG-be COMP sheep ART:DEF.SG s:3SG-cop farmer
nènè fọrẹ nà.
REL surpass RC

‘There was this town in which Sheep was the greatest farmer.’ \hfill (Atoyebi 2010: 278-279)
Exclusive-specific article

Specific referent:

(44) [èdèda òbèn] wo i-tò-ócín nènè e-te égbèn
father ART:EXSPEC.SG LOC.COP LOC-POSS:1PL-compound REL PROG-teach children
ábe ìkèwú nà
ART:DEF.PL Arabic.script RC
‘A (certain) man lives in our compound who teaches children the Arabic script.’

(Atoyebi 2010: 247)

Nonspecific referent:

(45) te-è-mín gba óró nènè e-mè-ni ka àyè sù épán óbòrò nà.
1PLS.-NEG-INCEP see person REL 3SGS.-NEG-want COMP 3SG.I have head good RCP
‘We have not seen a person who does not want to be fortunate.’

(Atoyebi 2010: 219)

Runyankore

Anaphoric article

(46) nya-0-baruha mu-a-gi-ha o-mu-shaija?
‘Have you given the said letter to the man?’

(Asiimwe 2014: 225)

Inclusive-specific article

(47) o-mu-piira gu-aa-baruka
ART:INSPEC-CL3-ball 3-PST.IMM-burst
‘The/a ball has burst.’

(Asiimwe 2014: 128)

Sandawe (Definite article)

Anaphoric referent:

(48) a. ùtwá lóóólò, nàŋgwé hàà bùrù-ki-àà l’úmá ts’õntò ...
long.ago INT cat and mouse-TOP-FOC earth small ...
‘A very long time ago, a cat and a mouse lived in a small world ...’

b. pàà bùrù-ŋg-àà ká?á friend ...
CONJ.3 mouse-ART:DEF-FOC that friend ...
‘And the mouse said: my friend ...’

(Steeman 2011: 74)
Contextually unique referent:

(49) ŋǁòó ŋàtə-ŋ swàkú-wàʔôŋ ǀàn=nì=i child leg.poss-art:DEF swell.poss-pl-inf see=pol=2sg
‘Have you seen the swellings of the legs of the child?’ (Steeman 2011: 246)

Sheko (Definite article)

Anaphoric referent:

(50) só ōtì há-s-tà ŋ=šà-fin-á-mə há=ge-ŋ
up.there cow prox.m-loc exist-rel-loc 1sg=arrive.nv-descend-put-IRR-sti 3ms=say ds
ōyt-ŋ datà ū=sâk-ŋ
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 …” ’ (Hellenthal 2010: 144)

Contextually unique referent:

(51) a. kâdū yi=haay-ǹtə kâdū ìʃ=túũtʂ-á-m.
three s:3.f=spend.night-cond three 3pl=knot-put-IRR
‘When (the moon) appeared in the third night, they made a third knot.’

b. súkú guru ìʃ=túũtʂ-gé-t=ìʃi ìʃ=túũtʂ-gé-t=ìʃi ját-ǹ-ʃ
rope only 3pl=knot-say-ss=3pl knot-say-ss=3pl maize-art:DEF-m
gâar-ki-bâåstə
bear.fruit-exist-while
‘Knotting and knotting that rope while the maize was ripening, …’
(Hellenthal 2010: 110)

Supyire (Definite article)

Anaphoric referent:

(52) a. lira a míì tú-ŋi ta u mpyi ná pwunm-pole emph.cl3.sg perf my father-art:DEF.cl1.sg get cl1.sg be.pst with dog-male è.
with
‘At that time my father had a male dog.’
b. uru pwùn-ŋa à pyi a sini mìi tú-ŋi
   EMPH.CL1.SG dog. ART:DEF.CL1.SG PERF PST PERF lie.down my father. ART:DEF.CL1.SG
   ba-gé ɲwò-gé na.
   house-ART:DEF.CL2.SG mouth-ART:DEF.CL2.SG at
   ‘This dog had lain down at the door of my father’s house.’

(Carlson 1994: 617)

Contextually unique referent:

(53) na jò u ɲyε na ci-ré pààn-ni ye?
   that who CL1S be PROG tree-ART:DEF.CL4 chop-IPFV Q
   ‘Who is it that is chopping the trees?’

B.2 Eurasia

Aghul

Indefinite article

Specific referent:

(54) aχpːa me χinebi qaːq’.a-j-e [sa q’isə] hate tūhûr hûr.i-ʔ uč.i-l
    then DEM Khinebi tell.IPFV-CV-COP ART:INDEF story DEM Tūkûr village-INREFL-SUPER
    ałcarx.u-nde
    SUPER-LAT.fall.PFY-PTCP
    ‘Then this Khinebi is telling us a story about an incident, that happened to him in that
    village of Tūkûr.’

Aghul (Ganenkov et al. 2009: 6)

Nonspecific referent:

(55) kupecːa p.u-naj, zun me ze ruš zun e-s-tawa šuw.a-s
    merchant.ERG say.PFY-AOR I DEM my daughter I give.INF-INF-COP:NEG husband-DAT
    [sa hunar] χìli-ʔ a-dawa-t:i-s.
    ART:INDEF skill hand-IN be.IN-NEG-S-DAT
    ‘The merchand said: “I will marry my daughter only to someone who can do something
    special.” ’

(Maisak 2014: 55)

1Original Russian translation: “Потом этот Хинеби рассказывает нам про один случай, который
произошел с ним в этом селе Тюхюр.”

2Original Russian translation “Купец сказал: «Я свою дочь выдам замуж только за того, кто умеет делать
что-нибудь необыкновенное».”
Albanian

Definite article

Contextually unique referent:

(56) a. Në atë shkollë ka rreth dyqind nxënës.
    in dem school there.are round two.hundred student
    ‘This school has about two hundred students.’
    
    b. Djem-të e vajza-t mësojnë me zell.
        boy-Art:Def.pl and girl-Art:Def.pl learn with diligence
        ‘The boys and girls are learning with diligence.’
        (Buchholz & Fiedler 1987: 235)

Anaphoric referent:

(57) a. Na ishte njëhërë një mbret plak.
    us was once Art:Indef king old
    ‘There was once an old king.’
    
    b. Mbret-i u sëmur.
        king-Art:Def.m 3pl sicken
        ‘The king got sick.’
        (Buchholz & Fiedler 1987: 235)

Indefinite article

Specific referent:

(58) Tashti kishte në shpinë [një brengë të vazhdueshme].
    now had in spine Art:Indef sorrow nmod continued
    ‘Now he had a continuous worry on his back.’
    (Newmark et al. 1982: 150)

---

3 Original German translation: “In dieser Schule sind rund zweihundert Schüler.”
4 Original German translation: “Die Jungen und Mädchen lernen fleißig.”
5 Original German translation: “Es war einmal ein alter König.”
6 Original German translation: “Der König wurde krank.”
Nonspecific referent:

(59) Po tẽ mẽ dhuronin nesër [një fjalor], pasnesër do tẽ if s:2sg o:1sg give tomorrow ART:INDEF dictionary day.after.tomorrow IRR PREP filloja tẽ mësoj shqip.
begin PREP learn Albanian

‘If I was given a dictionary tomorrow, I would start to learn Albanian the day after tomorrow.’

(Buchholz & Fiedler 1987: 145)

Ao

Anaphoric article

(60) a. tɔ̃-ɔr tɔ̃mhnak nɔ a-hnáʔ tʃu a-jim nɔ hɔn-ɔr tʃuwa
thus-SEQ INAL-son.in.law AG AL-fish DIST AL-village ALL take-SEQ emerge.PST
‘And then, the son-in-law took the fish and returned to the village.’

b. ...tɔ̀-li-kula [a-hnáʔ sa] hman-tʃak-ar tʃu ku jip
thus-BE-CIRC AL-fish ART:ANA set-RES-SEQ just 3DU AL-DIST LOC lie.down.PST
‘That being the case, he just set the (aforementioned) fish down and they copulated there.’

(Coupe 2007: 461-462)

Armenian

Definite article

Contextually unique:

(61) č’nayac [ɛ’urt elanak-i-n] na bavakanin t’et’ew ēr hagn-v-ac
PREP cold weather-DAT-ART:DEF he quite light was dress-REFL-PTCP:RES
‘Despite of the cold weather he was dressed quite lightly.’

(Dum-Tragut 2009: 710)

Anaphoric

(62) a. Milena-n yerek šune arel.
Milena-ART:DEF yesterday dog buy.PST
‘Milena bought a dog yesterday.’

---

7Original German translation: “Schenkte man mir ein Wörterbuch, würde ich übermorgen beginnen, Albanisch zu lernen.”
b. yerp šunê aračin ank’am tesav ir tunê, vaxecav. when dog.ART:DEF first time see.PST POSS:3SG house.ART:DEF get.scared.PST ‘When the dog saw her house for the first time, it got scared.’

(prim. data)

**Indefinite article**

Specific referent:

(63) mi tar-un na p'akanagorc darj-av [mi mec ART:INDEF year-LOC he.NOM locksmith.NOM become-AOR.3SG ART:INDEF big gorcaran-un] factory-LOC ‘In (the course of) one year he became a locksmith in a big factory.’

(Dum-Tragut 2009: 106)

Nonspecific referent:

(64) Inj [mi at'oř] tv-ek’! 1.DAT ART:INDEF chair give-IMP.2PL ‘Give me a (any) chair.’

(Dum-Tragut 2009: 107)

**Bantawa (anaphoric article)**

(65) a. ɨktet rajkumar-ʔa dor-u, ŋ raja-ʔo i-cha-ʔa dor-u, ... one prince-ERG ask-P:3 yes king-GEN poss:3SG-child-FOC ask-P:3 ... ‘One prince came to ask for her, for the king’s daughter, …’

b. …tə [moko rajkumar-da] khat-ma-ŋa chunt-a-ŋ-a-heda. ...but ART:ANA prince-LOC go-INF-EMPH refuse-PST-PROG-PST-SIM.P ...but she refused to go with the prince.’

(Doornenbal 2009: 96)

**Barwar, Neo-Aramaic (Exclusive-specific article)**

Specific referent:

(66) a. íθwa xa-málka. there ART:PRET-king ‘There was a king.’

b. áwwa málka xzéle xa-xålma. this king dreamt ART:PRET-dream ‘This king had a dream.’

(Khan 2008: 1492-1493)
Nonspecific referent:

(67)  héti skina!
   bring.IMPER knife!
   ‘Bring a (any) knife!’  (Khan 2008: 248)

Basque

Inclusive-specific article

Contextually unique referent:

(68)  nire lagun-a gure herri-ko alkate-a da.
   my friend.ART:INSPEC our town-rel mayor-ART:INSPEC is
   ‘My friend is the mayor of our town.’  (Hualde & Ortiz de Urbina 2003: 180)

Anaphoric referent:

(69)  a.  Uste dut euskalki-an ezin d-ela idatz-i, baserritarr-en
   think AUX:3PL.ERG dialect-LOC cannot AUX-COMP write-PRF peasant-GEN
   kontu-a d-ela.
   thing-ART:INSPEC is-COMP
   ‘They thing that one cannot write in the dialect, that it is a thing of peasants.’

   b.  Ni-k erakuts-i nahi iza-n dut euskalki-ak edozer-tara-ko
   I-ERG show-PFV want be-PFV AUX:1SG.ERG dialect-ART:INSPEC.PL anything-ALL-REL
   balio du-ela.
   value has-COMP
   ‘I have tried to show that the dialect can be used for anything.’
   (Hualde & Ortiz de Urbina 2003: 907)

Specific referent:

(70)  [Kotxe berri-a] erosi dut.
   car new-ART:INSPEC buy AUX
   ‘I’ve bought a new car.’  (Hualde & Ortiz de Urbina 2003: 119)

Presentational article

Specific referent:

(71)  Bilbon [etxe bat] erosi behar dut
   Bilbao.LOC house ART:PRES buy need AUX
   ‘I need to buy a house in Bilbao.’  (Hualde & Ortiz de Urbina 2003: 694)
Nonspecific referent:

(72)  zergatik ez erregutu hari [bitxi bat-zuk] utz zieazkion?  
why not beg her jewel ART:INDEF-PL lend AUX.S.COMP  
‘Why not beg her to lend her some jewels?’  
(Hualde & Ortiz de Urbina 2003: 503)

Bonan

Indefinite article

Specific referent:

(73)  atcan manda [mehtox tɔxax=dɔ] oɔ-tɔo  
3SG 1SG.LOC flower white=ART:INDEF give-IPFV.O  
‘He gave me a white flower.’  
(Fried 2010: 146)

Nonspecific referent:

(74)  dɔzɔma [htɔxṭa χɔnɔ=dɔ] ap-ku taraŋ or-tɔo  
Droma bicycle new=ART:INDEF take=IPFV-NMLZ desire come-IPFV.O  
‘Droma wants to buy a new bike.’  
(Fried 2010: 82)

Domari

Indefinite article

Specific referent:

(75)  pandʒi adḏir-ahr-i kury-ak  
3SG rent-INTR-PRED.SG house=ART:INDEF  
‘He is renting a house.’  
(Matras 2012: 105)

Nonspecific referent:

(76)  ū n-h-e’ wāši-s wala qirš-ak aha kurdī.  
and NEG-is-NEG with-3SG no penny=ART:INDEF this.m Kurd  
‘And this Kurd doesn’t have a penny.’  
(Matras 2012: 104)
Das heiße Wetter in Nordeuropa sorgt für eine erhöhte Waldbrandgefahr. 'The hot weather in Northern Europe has led to an increased risk of wildfires.' (prim. data)

Wir haben gestern eine Melone gekauft; wir müssen [die Melone] heute essen. 'We bought a melon yesterday; we need to eat the melon today.' (prim. data)

Ich habe [ein Geschenk] für dich. 'I have a present for you.' (prim. data)

Ich brauche schnell [einen Stift] zum Unterschreiben. 'I quickly need a pen to sign.' (prim. data)
Hakhun Tangsa

Definite article

Contextually unique referent:

(81) [ʒube ty haʔsuŋ bə] mama, [miʔ haʔsuŋ bə] mama, mama ny tuŋ ghost family village ART:DEF separate man village ART:DEF separate separate LOC sit ta?
  PST:3
  'The ghost village is separate, and the men’s village is separate. They lived separately.'
  (Boro 2017: 603)

Anaphoric referent:

(82) a. haʔsúŋ mɤ̂-theʔ iróny təhjaʔ rwépe village cl-one there girl young
  'There was a young girl in a village.'

b. [təhjaʔ rwéņe bə] rî-tî-nà dv t-aʔ girl young ART:DEF sky-person-f COP PST-3
  'The young girl was a sky lady.'
  (Boro 2017: 228)

Hungarian

Definite article

Contextually unique referent:

  'The two year Finnish experiment of the introduction of the minimum wage did not fulfill the expectations.'
  (prim. data)

Anaphoric referent:

  'They opened a new museum in Budapest.'

b. Elakar-t-unk menni, de zárva volt [a múzeum]. want-PST1PL go.INF but closed was ART:DEF museum
  'We wanted to go (there), but the museum was closed.'
  (prim. data)
Indefinite article

Specific referent:

(85) Vett-em [egy új kerékpar-ot]; már mentem is az íródába buy.pst-s:1sg art:indef new bicycle-acc already go.pst1sg emph art:def office-ill vele.
3sg.instr
‘I bought a new bike; I already went to the office on it.’

Nonspecific referent:

(86) Kellene nekem egy új kerékpar, de még nem talál-t-am egy-et ami need.subj 1sg.dat art:indef new bicycle but yet neg find-pst-1sg one-acc rel tetszene nekem.
please.subj 1sg.dat
‘I need a new bike, but I haven’t found one yet that I like.’

Irish

Definite article

Contextually unique referent:

(87) Bhain mé taitneamh as an gceolchoirm ach amháin go raibh [an halla] did 1sg enjoy from art:def concert but one comp did art:def hall fuar.
cold
‘I enjoyed the concert except that the hall was cold.’

Anaphoric referent:

(88) A: Tá mé anseo don agallamh.
cop 1sg here for.art:def interview
‘I am here for the interview.’

B: [Don agallamh]! Ach nil na hagallaimh ar siúl go dtí for.art:def interview but no art:def prep interview.pl walk comp until maidín amárach.
morning tomorrow
‘For the interview! But the interviews aren’t taking place until tomorrow morning.’

https://www.rte.ie/easyirish/aonad1.html
Kharia

Presentational article

Specific referent:

(89) souʔb se maha beʔt=ɖom simɾa, [mop maɾa=te] daʔ kuy=oʔ.
   all ABL big son=poss:3 Simra ART:PRES cave=ABL water find=pst
   ‘His eldest son, Simra, found water in a cave.’ (Peterson 2011: 143)

Nonspecific referent:

(90) akhaʔ kinir=te [mop bacha] tol daʔ=siʔ=ta.
   deep forest=ABL ART:PRES calf tie PASS=pst=mid.prs
   ‘A calf is (regularly) tied in the deep forest.’ (Peterson 2011: 251)

Kurtöp

Indefinite article

Specific referent:

(91) woye khwe=gi zur=na thang [pang cingu-la the] nà
dem:up water=gen edge=loc field open.space small-idz ART:INDEF COP.EXIST.MIR
   ‘At the edge of the water down there, there is a small, open plane.’ (Hyslop 2011: 283)

Nonspecific referent:

(92) sha-the zu wotor zu-male mutna
   meat ART:INDEF eat like this eat-nmlz:irr COP.EXIST.NEG.MIR
   ‘(We) didn’t have a piece of meat to eat.’ (Hyslop 2011: 612)

Lepcha

Definite article

Contextually unique:

(93) ?ádo-sá ?ányúʔ-re satet ?átim go!
   2s.OBL-GEN ear ART:DEF how.much big be
   ‘Your ears are so big!’ (Plaisir 2006: 50)
Anaphoric:

(94) a. hu múng go-nu-re hudo-m-re **holmúng** yang 3sg devil be-ABL-ART:DEF 3SG.OBL-DAT-ART:DEF mountain.devil thus li-wám-bú. say-PROG-FACT
‘It is a devil, so we call it devil of the mountains.’

be.friendly-ABL-PROG
‘The Mountain Devil is very friendly to us Lepcha.’

(Plaisir 2006: 142)

Palula

Anaphoric article

(95) a. be eetáa yhéi-il-a ta hiimaál čhinj-i asaám 1PL.NOM there.REM come-PFV-MPL SUB glacier strike-CV 1PL.ACC take.away.PFV-MPL híṛ-a bhun áa lhaáṣṭ zhay-i de
‘When we reached there, an avalanche git and swept us away, to a flat place below.’

b. [se lháṣṭ-i-wée] de-i asaám be dharíit-a ART:DEF plain-obl-into give-CV 1PL.ACC 1PL.NOM remain.PFV-MPL
‘We were carried away to that flat place and came to rest.’ (Liljegren 2016: 429)

Exclusive-specific article

Specific referent:

(96) bhuná [áa ɡíri] heensil-i below ART:EXSPEC rock stay.PFV-F
‘Down below there was a big rock.’ (Liljegren 2016: 309)

Nonspecific referent:

(97) **muloó** díi yéeř kráam na bháan-u mullah from without work NEG become.PRS-M.SG
‘Without a mullah the work is not being done.’ (Liljegren 2016: 332)
Rajbanshi

Exclusive-specific article

Specific referent:

(98) [ek-ta maṣṭar] raha-e
   ART:PRES-CL teacher be-PST.HAB-3.
   ‘There was once a teacher.’
   (Wilde 2008: 72)

Nonspecific referent:

(99) tui jahaj dekh-ic-it?
   2SG aeroplane see-PERF-2SG
   ‘Have you seen an aeroplane?’
   (Wilde 2008: 81)

Sri Lanka Malay

Indefinite article

Specific referent:

(100) Sithu=ka, [hathu 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)

Nonspecific referent:

(101) 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)

Tamil

Indefinite article

Specific referent:

(102) kumaar [oru peṭṭ-y-ai] vaaṅk-in-aan.
   Kumar ART:INDEF box-ACC buy-PST-3SG.M
   ‘Kumar bought a box.’
   (Lehmann 1993: 29)
Nonspecific referent:

(103) kumaar [oru maappillai-y-ai-t] teetu-kir-aan.
Kumar ART:INDEF son.in.law-ACC look.for-PRES-3S.M
\(\text{`Kumar is looking for a son-in-law.'}\) \((\text{Lehmann 1993: 29})\)

**Turkish**

**Indefinite article**

Specific referent:

(104) Kahvaltidan sonra [bir karpuz] yedim.
breakfast after ART:INDEF watermelon ate.1sg
\(\text{`I ate a watermelon after breakfast.'} \) \((\text{Erguvanli & Zimmer 1994: 548})\)

Nonspecific referent:

(105) [Bir doktor] arıyorum.
ART:INDEF doctor seeks.PRES.1SG
\(\text{`I’m looking for a doctor (any doctor).'} \) \((\text{Erguvanli & Zimmer 1994: 548})\)

**B.3 Papunesia**

**Abui (Recognitional article)**

(106) ya mi ba it-i [tipai lil-a nu] he-tok-u ba ong
water take LNK lie.on-PFV iron hot-BE.AT ART:RECOG 3-LOC-drop-PERF LNK make
palak-n-a
cold-see.COMPL-DUR
\(\text{`Take some water and pour it on that hot iron so that it cools down.'} \) \((\text{Kratochvil 2007: 307})\)

**Agta**

**Definite article**

Anaphoric referent:

(107) 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
\(\text{`We also understand how to swidden farm here, here in Bolos.'} \)
b. kalan-an=tam=bi=dan hito [i mag-sikaw].
copy-loc=1pl.in.gen=also=already there \texttt{ART:DEF} \texttt{ACT-swidden}

‘We can all copy the swidden farming.’

\textit{(Robinson 2008: 324)}

\begin{itemize}
  \item \textbf{Contextually unique referent:}
  \begin{itemize}
    \item (108) awan=bi ito i-unnuy=pala [hi dinom=a]
      neg=also it theme-drift=still \texttt{ART:DEF} \texttt{OBL} water=\texttt{ART:EXSPEC}
      ‘It still isn’t sent downstream on the water.’

      \textit{(Robinson 2008: 93)}
  \end{itemize}

  \item \textbf{Exclusive-specific article}
  \begin{itemize}
    \item Specific referent:
      \begin{itemize}
        \item (109) mag-hen=kano ha rukab mangilen=a
          \texttt{AV-live=HSY} \texttt{OBL} cave ghost=\texttt{ART:EXSPEC}
          ‘They say a ghost was living in a small cave’

          \textit{(Robinson 2008: 94)}
      \end{itemize}

    \item Nonspecific referent:
      \begin{itemize}
        \item (110) awan=bi a nag-ga-galgal i ulitaw=a ha ogsa
          neg=also \texttt{LNK} \texttt{COMPL.AV-MULT} \texttt{SLICE} \texttt{ART:DEF} young-man=\texttt{ART:EXSPEC} \texttt{OBL} deer
          ‘The young man has never butchered a deer.’

          \textit{(Robinson 2008: 96)}
      \end{itemize}
  \end{itemize}

  \item \textbf{Bajau (Definite article)}
  \begin{itemize}
    \item Anaphoric referent:
      \begin{itemize}
        \item (111) a. Diam taun e uun di-kau’ gua.
          inside forest \texttt{ART:DEF} \texttt{EXIST} one-cl cave
          ‘Inside the forest there was a cave.’

        \item b. Kerabaw e posok en-diam [gua e].
          buffalo \texttt{ART:DEF} enter prep-inside cave \texttt{ART:DEF}
          ‘The buffalo entered into the cave.’

          \textit{(Miller 2007: 441-442)}
      \end{itemize}
  \end{itemize}
\end{itemize}
Contextually unique referent:

(112) Masa e, [anak kerabaw siari e] boi kawin engko’ anak rojo
time ART:DEF child buffalo youngest ART:DEF COMPL marry PREP child king
siari pan mandi me en-suang.
youngest TOP bathe there PREP-river
'That that time, the youngest daughter of the buffalo, who had married the youngest
prince, was bathing in the river.' (Miller 2007: 448)

Biak

Definite article

Anaphoric referent:

(113) 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.'

b. [mankroder anya] ma romá-nkun 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)

Contextually unique referent:

(114) Sko-ra ro [sorn anya] fadu-ri ...
3PC-go LOC deep.sea ART:DEF middle-poss:SG ...
'They arrived in the middle of the deep sea ...' (van den Heuvel 2006: 429-430)

Exclusive-specific article

(115) 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.” (van den Heuvel 2006: 432)
Nonspecific article

(116) sansun vye=na na-is-ya voz, 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.’ (van den Heuvel 2006: 400)

Bilua (Presentational article)

Specific referent:

(117) a. ...se ta ke ere=v=e [kala niabara]. 3pl top 3pl make=3sg.m.o=rmp art:pres.m.sg war.canoe ‘They made a war canoe.’
b. Ne=a niabara ta a=daite Böbe=vo. prox.sg.m=lig war.canoe top 1sg=grand.parent Boxe=3sg.m ‘This war canoe was my grandfather Boxe’s.’ (Obata 2003: 79)

Nonspecific referent:

(118) ...enge ta nge=q=ai zari=a tu k=ov=o [kama ...1pl.excl top 1pl.excl=3sg.f.o=val want=pres irr 3sg.f.o=get=nom art:pres.sg.f uri=a=ma saev=o] ... good=lig=3sg.f survive=nom ‘...we want to get a good life...’ (Obata 2003: 93)

Bunaq (Anaphoric article)

(119) 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)
Komnzo

Anaphoric article

(120) a. bthan kabe fthé fenz yona-si bänemr zrethkhäfth
    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=INST middle
    ‘...the sorcerer sets off to go to the grave yard, to the grave house. He goes secretly
    in the middle of the night.’
    (Döhler 2018: 408-409)

Presentational article

Specific referent:

(121) komnzo zena bobo rä ane kar we [nä fof
    still today MED.ABL be.NPST.IPFV.S:3SG.F ART:ANA village also ART:PRES EMPH
    rä trik-si] kar fof.
    be.NPST.IPFV.S:3SG.F tell-NMLZ village EMPH
    ‘This place is still there and there is also a story about that place.’
    (Döhler 2018: 372)

Nonspecific referent:

(122) zokwasi=mär ńafiyo kwa keke [nä zokwasi].
    word=PRIV make.PST.IPFV.S:3SG NEG ART:PRES words
    ‘He was speechless …no words whatsoever.’
    (Döhler 2018: 104)

Lavukaleve

Definite article

Anaphoric referent:

(123) a. meo vo-ha fi ma-me.
    tuna.PL 3PL:O-clear 3SG.N.FOC S:3PL-HAB
    ‘They were catching bonito.’
   tunapl ART:DEF.PL o:3PL-SUB-feed-ANT s:3PL-go.seawards
   ‘When the bonito were feeding, they went seawards.’

(Terrill 2003: 92)

Contextually unique referent:

(124) vela-nun [lai ga] hoa-re e-e-vo-ge ini
godur rain ART:DEF.SG.N poke.through-INF o:3SG-SUB-come-ANT VENT
ukuea-re lo.
come.close-INF finish
   ‘It went on, and the rain started falling, it up and came close.’
   (Terrill 2003: 71)

Recognitional article

(125) 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]
   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)

Indefinite article

Specific referent:

(126) Mima hona hau e’rau oiga enata hano [gu
   Mima MOD.PROX.SG.M go.shorewards fall other.NTRL.SG.N that.moment then wave
   ro]
   hale-re vo-a.
   ART:PRES.SG.F break-INF come-SG.F
   ‘At that very moment that this Mima went shorewards, a wave broke.’
   (Terrill 2003: 80)

Nonspecific referent:

(127) Ngali ra ba kini [foe roa] a-kula
   friend coconut.crab let’s.go VENT pig ART:PRES.SG.M o:3SG.M-run.after
   a-me-foi hide o-re.
   o:3SG.M-1DU.IN-hold thus s:3SG-say
   ‘“Friend coconut crab, let’s go catch a pig”, he said.’
   (Terrill 2003: 79)

---

The marker glossed as ventive here is originally glossed as “action marker” in Terrill (2003), and is described as referring to motion and action, as in “go and V” or “up and V” (Terrill 2003: 70).
Maori

Inclusive-specific article

Anaphoric referent:

(128) a. Ko Wairangi te tangata whakamutunga ki te tomo i TOP Wairangi ART:INSPEC man last to ART:INSPEC enter at te whare ...
ART:INSPEC house
'Wairangi was the last man to enter the house ...'

b. Kaatahi ka rere a Wairangi ki te tute i [te whare], kore then PST leap PROP Wairangi to ART:INSPEC shove do ART:INSPEC house NEG rawa i ngaoko.
EMPH PST stir
'Then Wairangi threw his weight against the side of the house, but it didn't yield in the slightest.'

(Bauer 1993: 153)

Contextually unique referent:

(129) Ka whakamiiharo anoo a Rewi ki [te kaimahi] whakahaere i PST wonder.at again PROP Rewi to [ART:INSPEC worker] operate do te miihini uta i ngaa raakau ART:INSPEC machine load do ART:INSPEC.PL tree
'Rewi marvelled again at the worker operating the machine loading the trees.'

(Bauer 1993: 460)

Specific referent:

(130) Ka takai-a, ka kawe-a, ka whaka-iri-a ki runga ki [te kauere]. PST wrap-PASS PST carry-PASS PST CAUS-hang-PASS to top to ART:INSPEC puriri
'They wrapped him up and took him and suspended him in a puriri tree.'

(Bauer 1993: 444)

Presentational article

Specific referent:

(131) Ka noho ia i [teetahi raakau] e noho nei he taangata i raro.
PST sit 3SG at ART:PRES.SG tree A sit PROX ART:INSPEC people at underneath
'He sat in a tree under which some people were sitting.'

(Bauer 1993: 55)
**Nonspecific referent:**

(132) Kaare anoo [teetahi raururu koohahi] kia paa.  
\[\text{NEG again ART:PRES problem one subj arise}\]  
Not a single problem had arisen.  
\[(\text{Bauer 1993: 297})\]

**Menya (Definite article)**

**Anaphoric referent:**

(133) I-ät-äng=ŋgaŋi qa ymekā tänā e-ät-in i=ŋgaŋi qa qe=qā  
\[\text{that-ipfv-ds time 2DU child near be-ss-2DU that=time 2DU 2DU=poss}\]  
\[\text{ymeqā=i u ä-w-mātuqā-iny=ŋqā=nā}\]  
\[\text{child=ART:DEF=LOC ASSERT-3-show-2DU=for=FOC}\]  
‘So you two are going to have children and then you are going to teach your children.’  
\[(\text{Whitehead 2006: 257})\]

**Contextually unique referent:**

(134) im ä-pma-njāqān=aqe hiunji quemisqā ayā he-q-ąnga, [ynŋā  
\[\text{that ASSERT-be-DUR-ss:1PL midday noon just be-pfv-ds-time bird}\]  
\[\text{ä-pātnū-k-qāu=i qa ä-mā-mi-mā}\]  
\[\text{ASSERT-shoot-pst-1PL=ART:DEF bag ASSERT-get-put.in-ser}\]  
‘We remained in there and when it was high noon we gathered the birds we caught into string bags.’  
\[(\text{Whitehead 2006: 244-245})\]

**Oksapmin**

**Definite article**

**Anaphoric referent:**

(135) a. go tap=xē pat=d=a m-p-n-gopa=li?  
\[\text{2SG pig=foc stay.ipfv.sg=q=empf o:prox-tell-pfv-evid.rempst.pl=report}\]  
‘Do you own a pig?’  

b. ...jexe [tap jox] su-ti-pa=li.  
...then pig ART:DEF kill-pfv-fact.rempst.pl=report  
‘Then they killed the pig.’  
\[(\text{Loughnane 2009: 499})\]
Contextually unique referent:

(136) tom jox lum p-d-m edi-pla=o [ake jox]
  water ART:DEF a.lot CAUS-eat-SEQ stay.PFV-REMREMUT.SG=QUOT stomach ART:DEF
  ox=o tom=wi x-ti-plex=xejox n-pli-nuŋ.
  2SG.M=QUOT water=only be-PFV.HODREMUT.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.’
  (Loughnane 2009: 480)

Recognitional article

(137) gin i ml-sa jæxe 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 …’
  (Loughnane 2009: 124)

Indefinite article

Specific referent:

(138) jæxe nox øpi-s kip jox øpi-s kol jox [xan pæsel tit]
  then 1SG come-SEQ road ART:DEF come-SEQ arrive.PRS.SG TOP man old ART:INDEF
  øpli-pat-nuŋ
  come-IPFV.SG=EVID.HODPST.SG
  ‘When I came to the road, (I saw that) an old man was coming along.’
  (Loughnane 2009: 486)

Nonspecific referent:

(139) em go [dup tit] n-a-xu-ti-n=a!
  mother.POSS:1 2S bow ART:INDEF O:1-BEN-twirl-PFV-IMPER=EMPH
  ‘Mum, twist a bow for me!’
  (Loughnane 2009: 243)

Papuan Malay (Definite article)

Anaphoric referent:

(140) a. …pas ketemu deng sa pu nene, nene, trus kaka
  …precisely meet with 1SG POSS grandmother grandmother next O:SB
  laki-laki, mama-tua pu ana.
  RDP-HUSBAND aunt POSS child
  ‘[I passed by and reached the village market there, I was sitting, standing there.] right
  then I met my grandmother, grandmother, and then my older brother, aunt’s child.’
b. baru [nene de] mulay tanya saya, de blang ...
and.then grandmother ART:DEF start ask 1sg 3sg say ...
‘And then grandmother started asking me, she said ...’

(Kluge 2017: 355)

Contextually unique referent:

(141) macang kalo [bapa de] kasi bona ini, a, nanti ...
variety if father ART:DEF give girl DEM:PROX ah very.soon
‘[About bride-price children:] for example, if (my) husband gives this (our) girl (to our relatives), ah, later ...’

(Kluge 2017: 354)

Rapa Nui

Referential article

Contextually unique referent:

(142) I hāŋai era i nunui era, te aŋa o tū rū’au māmā era he
PFV raise DIST PFV PL.big DIST ART:REF work of DEM old.woman mother DIST PRED
kā i [te ‘umu] paūrō te mahana.
kindle ACC ART:REF earth.oven every ART:REF day
‘When they had raised them and they had grown up, what the old mother did was cooking
food in the earth oven every day.’

(Kieviet 2017: 579-580)

Anaphoric referent:

(143) ki a kōrua [ki te nu’u] hakarojo mai ī 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 ...’

(Kieviet 2017: 46)

Specific referent:

(144) ī te noho iŋa tuai era ’ā [te taŋata 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.’

(Kieviet 2017: 238)
Nonspecific referent:

(145) ...mo ai o te moni mo ho’o mai i [te haraoa].
...for exist of ART:REF money for trade hither ACC ART:REF bread
‘(He sells food) in order to have money to buy bread.’ (Kieviet 2017: 238)

Siar Lak

Inclusive-specific article

Contextually unique referent:

(146) A rak al an, ap al an o-n [ep wang] né.
1SG want 1SG.POT go and 1SG.POT go OBL-3SG ART:INSPEC canoe this
‘I want to go, and I want to go in this canoe.’ (Rowe 2005: 25)

Anaphoric referent:

(147) 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 …’

Specific referent:

(148) 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.’ (Rowe 2005: 50)

Nonspecific article

(149) 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 …’ (Rowe 2005: 18)
Sunwadia

Referential article

Anaphoric referent:

(150) a. Ete sae le=woña-na, saqे na=qona kat vene min look lift LOC=top-poss:3sg find ART:REF:ACC=sarcophagus conj shoot GOAL a=no-n tīqaña.
   ART:REF=cl:poss-poss:3sg arrow ‘He looked up towards [its] top, noticed a sarcophagus, and shot an arrow.’9

b. Vene sae le=woña-na, awe dan a=qona.
   shoot lift LOC=top-poss:3sg miss abl ART:REF=sarcophagus ‘He shot into the air towards the top, and missed the sarcophagus.’10
   (Henri 2011: 439)

Contextually unique referent:

(151) Nau=ete [na=no-ra naba], i=wa=tarisa min s:3sg=look ART:REF:ACC=cl:poss-poss:3pl number s:3sg=frame be.equal GOAL a=resista ...
   ART:REF=register ...
   ‘I checked their number, and if it were identical with the one of the register …’11
   (Henri 2011: 126)

Specific referent:

(152) Kami=roño na=tatua sumai ti=sañav na=gatama.
   s:1pl.ex=hear ART:REF:ACC=man come aor.3sg=open ART:REF:ACC=door ‘We were sleeping, and we heard a man come and open the door.’12 (Henri 2011: 121)

Nonspecific referent:

(153) I=u=saro-ra na tatua be gani-ra.
   s:3sg=perf:res=search-o:3pl ART:REF:ACC man purp eat-o:3pl ‘He looked for men to eat.’13
   (Henri 2011: 122)

---

9 Original French translation: “Il leva les yeux vers la cime, aperçut un sarcophage, et décocha une flèche.”
10 Original French translation: “Il tira en l’air vers la cime, et rata le sarcophage.”
11 Original French translation: “Je vérifiais leur numéro, et s’il était identique à celui du registre, …”
12 Original French translation: “Nous dormions, et nous avons entendu un homme venir et ouvrir la porte.”
13 Original French translation: “Il cherchait des hommes à manger.”
Sye (Presentational article)

Specific referent:

(154) Kamli-ve ra [hai nu orog].
    1PL:EX.RECPST-go GOAL ART:INDEF river big
'We went to a big river.'

(Crowley 1998: 164)

Nonspecific referent:

    if ART:INDEF gust small 3SG.PST:HAB-come 3PL.PST:HAB-shout
    'If a small gust came, they would shout.'

(Crowley 1998: 270)

Teiwa

Anaphoric article

(156) a. ...ana’ maan si ki uwaad nuk yaa, bif ga’an tu’uk.
    ...long.time NEG SIM eagle big ART:PRET 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, ...'

(Klamer 2010: 433)

Presentational article

Specific referent:

(157) 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:PRET be drop
    'So that deer takes the child to a river bank and drops it.'

(Klamer 2010: 433)

Nonspecific referent:

(158) Naree maan, xoran si ma ha gi [ya’] siis nuk
    grandfather NEG thus SIM 2SG go kind.of.small.bamboo dry ART:PRET
    ga-uyan pin aria’, ...
    3SG-look.for hold arrive ...
    'Grandfather no, if so you go look for dry bamboo to bring here ...'

(Klamer 2010: 340)
Tongan

Inclusive-specific article

Anaphoric referent:

(159) a. Pea folofola 'a e 'Otua, ke maama mai: pea na’e maama. and HON: say ABS ART: INSPEC God that light DIR: us and PST light ‘And God said, Let there be light: and there was light.’

b. Pea 'afio 'e he 'Otua ki [he maama], kuo lelei ... and HON: see ERG ART: INSPEC God all ART: INSPEC light PERF good ... ‘And God saw the light, that it was good ...’

Contextually unique referent:

(160) 'Oku hiva 'a Sione 'i [he Siasi Uesiliana], PRES sing ABS John LOC ART: INSPEC Wesleyan Church ‘John sings in the Wesleyan Church.’ (Völkel 2010: 115)

Specific referent:

(161) ko 'eku lele mai pe 'o fakaa'u 'a [e 'ofa] mei he PRES POSS: 1 SG 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: INDEF SPEC Lolo 'a Halaevalu prepare remind ABS ART: INSPEC valu-ngofulu-ma-nima ta’u 'a e 'Afio na 58th 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.’ (Völkel 2010: 208)

Nonspecific article

Nonspecific referent:

(162) ha’u mo [ha afo] come with ART: INSPEC fishing line 'Bring a (any) fishing line.' Tongan (Churchward 1985: 25)
Urama

Definite article

Anaphoric referent:

(163) 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.’

(J. Brown et al. 2016: 22)

Contextually unique referent:

(164) [Nu moto=i] umu hiro-hia ka.
3sg house=ART:DEF dog many-very PRS
‘[In that village there lived a certain man …] At his home there were a lot of dogs.’

(J. Brown et al. 2016: 85-86)

Anaphoric article

(165) a. vloro ohu=i tabo kiaukia bomo gema=i ro go’ota=i ahiai climb top=ART:DEF LOC enough pig big=ART:DEF NOM coconut=ART:DEF cut ka. 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.’

(J. Brown et al. 2016: 89)
B.4 Australia

Arrernte (Anaphoric article)

(166) a. Lhere itere-ke-rl-irre-ke anteme kwele re pety-alpe-tyeme creek.bed side-DAT-REL-INCH-PST:COMPL now QUOT S:3SG come-back-PST:PROG kwatye iperte ikwere-werne ... water hole 3SG.DAT-ALL ... ‘She [his mother] had reached the side of the creek and was returning to the water-hole ...’

b. ...Kenhe kele [relhe nhenge] petye-anteme-alpe-rlenge, itwe anteme. ...but o.k. woman ART:ANA come-now-back-DS near now ‘...Then, his mother finally returned and neared him.’ (Wilkins 1989: 530, 541)

Bardi (Anaphoric article)

(167) a. Ginyinggo ingarrjarralana Giido injidar Ganbaliny. then they.ran Giido husband.and.wife Ganbaliny ‘Then Giido and Ganbaliny ran.’

b. Injalana arrjambala [jarr injidar irr Giido Ganbaliny]. he.saw their.tracks ART:ANA spouse they Giido Ganbaliny ‘He saw the tracks of the husband and wife Giido and Ganbaliny.’ (Bowern 2008: 85)

Bininj Kun-Wok (Recognitional article)

(168) Well sometimes barri-re-i gonda [garrigad na-wu] ... well sometimes 3.PST-go-PST.IPV there high.country M-ART:RECOG ‘When they’d go up into that high country ...(first mention of the high country after some lines discussing lowland activities)’ (Evans 2003: 297-298)

Diyari (Definite article)

Anaphoric referent:

(169) a. [nhulu kanku-yali] wata ngantya-yi ngamalha. ART:DEF.NF.ERG boy-ERG not want-PRES sit-ss ‘The boy didn’t want to stay.’

b. nhawu dalki-yi pulangu ya nhungkarni ngandrinhi. 3SG.NF.NOM disobey-PRES 3DU-LOC and 3SG.NF.DAT mother-LOC ‘He disobeyed them and his mother.’
c. [nhawu kanku] ngupaŋŋa mindi-rna.
   ART:DEF.SG.NF.NOM boy ahead run-PTCP
   ‘The boy ran ahead.’ (Austin 2017: 1-2)

Contextually unique referent:

(170) ngathu nhanha wilha nhayi-rna wara-yi yindra-rnanhi
   1SG.ERG ART:DEF.SG.F.ACC woman.ABS see-PTCP AUX-PRES CRY-1PFV.DS
   ‘I saw the woman who was crying.’ (Austin 2011: 232)

Gooniyandi (Recognitional article)

(171) [ginarndi yoowooloo] jijaggiddaa-nhi wambiggoowaari.
   ART:RECOG man speak.PROG.1PL-of him go.inside.PROG.3SG
   ‘The man who we’re talking about is going inside.’ (McGregor 1990: 438)

Guragone (Anaphoric article)

(172) Arrapu mi-nji lay, mi-nji [mu-garrapu moligiyi], moligiyi, yakala, and
   niyépu mu-garrapu, muntgomómortji njiwu-ba-ngarni mu-garrapu butj
   guwarr. long.ago
   ‘And which food hey, which (were) the ones I’m talking about, green plum, green plum,
   green plum [other type of green plums], he’s the one I’m talking about, red apple we used
to eat, those ones, long ago in the bush.’ (R. Green 1995: 68)

Jingulu (Anaphoric article)

(173) Nginda wurruku [kuna-ngka biningkurrung-gka].
   that.M 3PL-went ART:ANA-ALL billabong-ALL
   ‘They went to that lake [you know the one].’ (Pensalfini 1997: 232)

Mangarrayi (Anaphoric article)

(174) a. Ø-warawarama-ŋ landi mayawa mid ja-Ø-daya ṇaŋ-nawu-bayi
   A:3SG.P:3SG-shake-PST tree now tighten 3-A:3AG.P:3SG-AUX neck-ITS-FOC
   mayawa-ja
   NOW-EMPH
   ‘He [the boy] shook the tree, now tightens (the noose) around its [the goanna’s] neck,
now then!’
   child
   ‘ “Pull, let’s you and I pull together”, said the boy.’
   (Merlan 1989: 45)

Martuthunira (Anaphoric article)

(175)  a. Ngana nhula kanyara?
   who near.you man
   ‘Who is this man?’

b. Wantha-nyungu, mirta warnu ngayu nhawu-wayara yimpala-a-wa
   where-DWELL not ASSERT 1SG see-HABIT like.that-ACC-you.know
   man-ACC ART:ANA stranger really-now
   ‘He’s from somewhere else, I haven’t ever seen a man like that. This is a real stranger.’

c. Wanthala-nguru-lwa [nhiyaa kanyara]?
   where-ABL-ID ART:ANA man
   ‘Where’s he [this man] from exactly?’
   (Dench 1994: 132)

Wardaman (Anaphoric article)

Merlan (1994: 96) notes that example (176) “occurred in a story about a fight between brolga and emu, in which both had previously been introduced, but emu had not been mentioned for some time.”

(176) gumurrinji-marlang bagbag 0-bu-ndi yi-warli bag yi-bam darang
   emu-ART:ANA break 3-AUX-PST YI-arm-ABS break YI-head-ABS rump-ABS
   waryana.
   and.all
   ‘He broke the arm of that emu, broke his head and rump and all.’
   (Merlan 1994: 96)

Warrwa (Anaphoric article)

(177)  a. linyju-nma ngirrmurunguny-jina.
   policeman-ERG looked.3PL-3MIN.OBL
   ‘The policemen were looking for him.’
b. [kinya-na linyju], murunguna-yina.
   ART:ANA-ERG policeman looked.3SG-3MIN.OBL
   ‘So the policeman went looking for him.’
   (McGregor 1994: 61)

Wubuy (Anaphoric article)

(178) a. sa-ji-ruj alagala, ni-yan=dhawiwa:’ yaga ni=yama:’maga:
   IMM-LOC-LOC halfway 3SG.M-voice=ask.PST well 3SG.M=do.that.PST
   ‘A little further on, along his way, he asked (a snake), “Well? How much further?”’
   ,
   b. jujuj! wu=yama:’ an-uba-ni-yun ana-marn ana-lhiribala
   go CL=do.that.PST CL-ART:ANA-CL-ABS CL-snake CL-under
   ‘That snake, under the ground, said “Keep going!”’
   (Heath 1980: 19)

Yankuntjatjara (Recognitional article)

(179) munu nyaku-payi “waru kampa-nyi [kapi panya-ngka-manti-r]
   ADD see-CHAR fire.NOM burn-PRES water ART:RECOG-LOC-PROB-EXCIT
   ‘And they would see “There’s a fire burning, at that water(hole), you know the water(hole),
   most likely, by gee.”’
   (Goddard 1985: 54)

Yolŋu (Anaphoric article)

(180) ŋayi marrtji dhal’yu-n [unjhi be-ŋur-nydj a burumun’-ŋur-nydja dijarr]
   3SG go-1SG land-1SG ART:ANA INDEF-ABL-PROM island-LOC-PROM plane
   ‘it landed, the plane from the island (Bali) [earlier in the text, which was told at Gali-
   win’ku, Darwin airport was established as the location using the distal demonstrative.]’
   (Wilkinson 1991: 263)

Yuwaalaraay (Anaphoric article)

(181) a. giiruu nhama garigari gi-gila-nhi maadhaay-dji
   true.very 3SG afraid get-CTS-PST dog-ABL
   ‘He was afraid of the dog.’
   b. giiruu [ngam’ maadhaay-u] gawaa-nhi [ngam’ birralii-duu]l
   true.very ART:ANA dog-ERG chase-PST ART:ANA child-DIM
   ‘The dog chased him [the little boy].’
   (Giacon 2014: 191)
B.5 North America

Chatino

Presentational article

Specific referent:

(182) mn7q³ meřlo³¹ [skaf kto³] ka¹³
   see.compl Carmelo ART:PRES chicken yesterday
   'Carmelo saw a chicken yesterday.'

(McIntosh 2011: 106)

Nonspecific referent:

(183) s7i²("o) knya⁷³¹ kaʳ nu² mn7q²¹
   NEG deer be.PROG REL see.compl.2SG
   'You did not see a deer.'

(McIntosh 2011: 104)

Chimariko

Definite article

Anaphoric referent:

(184) a. h-iṭa-wiˀmut h-ičʰe-mda h-iṭa-md̬u
   3-hand-take 3-say.PROG POSS-hand-INSTR
   'He took his hand telling him (to go home), he led him by the hand.'

b. qʰomal uwa-m-aˀ č’imar-op?
   where go-DIR-Q person-ART:DEF
   'Where did that man go?'

(Jany 2009: 214)

Contextually unique referent:

(185) masunu h-uwa-kta-nhut šunuhull-ot.
   always 3-go-DIR-CONT old.woman-ART:DEF
   'The old woman (Mrs. Bussell) goes around all the time.'

(Jany 2009: 63)
Chocholtec (Presentational article)

Specific referent:

(186) Jorge chunda [naa kulucho].
    Jorge has ART:PRES donkey
    'Jorge has a donkey.'\(^{14}\)

In Campbell & Luna Villanueva (2011), there are no examples of a nonspecific referent.

Ch’ol

Definite article

Anaphoric referent:

(187) a. an\-ø=bi ji\-ni, ju\-\-tyikil wi\-\-n\-ik.
    E-B3=REP hm ART:INDEF-CL man
    ‘There was a man.’

b. no se chuki tyi ma i-k’el-b-e\-\-n\-ø.
    NEG know.1SG what PRFV away A3-see-APL-DT-B3
    ‘I do not know what he went to see.’

c. ts\-\-a=bi och-i\-\-ø tyi i-mali ch’e\-\-n [li wi\-\-n\-ik=i].
    PRFV=REP enter-IV-B3 PREP A3-inside cave ART:DEF man=FIN
    ‘The man went into the cave.’

      (Vázquez Alvarez 2011: 245-246)

Contextually unique referent:

(188) ts\-\-a=x lajm-i\-\-ø [li y-\-\-otyoty la=k-\-\-pi ál-ob-tyak=i].
    PRFV=already finish-IV-B3 ART:DEF A3-house PL.IN=A1-friends-PL3-PLINDF=FIN
    ‘Our friends’ houses were finished.’

      (Vázquez Alvarez 2011: 246)

Indefinite article

Specific referent:

(189) poj a\-\-n lonk-sa’ tyi [jum-p’e bolsaj]
    HON EXIST 1PL.EX pozol in ART:INDEF-CL:PEJ plastic.bag
    ‘We have our pozol in a plastic bag.’

      (Vázquez Alvarez 2011: 171)

\(^{14}\)Original Spanish translation: "Jorge tiene un burro."
Nonspecific referent:

a. tyi j-k’ajty-i-b-e [jum-p’e baso]  
Pfv s:1sg-ask-tr-appl-tr art:indef-cl:p’e glass  
‘I ask him for a glass.’  
(Vázquez Alvarez 2011: 187)

Chumash

Definite article

Anaphoric referent:

(190) a. he’si-’at’axatš ’an ts-kitwo ts-uniyəw lo’ka-nunašǝš.  
Dem:prox-man top 3-go.out 3-search.for art:Def-demon  
‘This man went out hunting the demon.’  
b. k’uwe ts-uštǝł.  
but 3-find  
‘And found him.’

c. kika=ts-’ip-us lo’ka-nunašǝš …  
sbo=3-tell-apl art:Def-demon …  
‘And the demon told him …’  
(Henry 2012: 436)

Contextually unique referent:

(191) lo’ka-x’anwa ’an ts-oxpot lo’ka-ts-owo.  
art:Def-woman top 3-untie art:Def-3-hair  
The woman untied her hair.”  
(Henry 2012: 443)

Indefinite article

Specific referent:

(192) ts-’ił-waš si-’at’axatš ’an ts-’ił si-ts-šatiwə.  
3-exist-pst art:indef-man rel 3-exist art:indef-3-spouse  
‘There was a man and he had a wife.’  
(Henry 2012: 435)

Nonspecific referent:

(193) k-suya-hin si-tštǝ’aniwaš.  
1-desid-have art:indef-dog  
‘I would like to have a dog.’  
(Henry 2012: 336)
Cupeño (Definite article)

Anaphoric referent:

(194) a. Mu=ku’ut hashi-peayax=ku’ut, pe-na’aqwa pe-yawichin. and=quot go=3sg=quot 3sg-daughter 3sg-take 'And it is said he went off, he took his daughter, it is said.'

b. Me=ku’ut [pe’=e pe-na’aqwa] pe-yka piyama mangin pe-neq. ... and=quot ART:DEF=FOC 3sg-daughter 3sg-behind always slowely 3sg-come ... 'And it is said that his daughter kept coming along slowly behind him, ...'

(Hill 2006: 504-505)

Contextually unique referent:

(195) Me [pe’ mulu’we-t] pe-tul-qa. and ART:DEF first-NPN 3sg-finish-PRS 'And the first one finishes.' (Hill 2006: 356)

Haida (Definite article)

Anaphoric referent:

(196) a. 7aajguusda-gee-raa-7isan ginntiigaas st’a.àwga-gaa-n. on.this.side-ART:DEF-in-too animals be.full-EVID-PST 'The other side too was full of animals.'

b. 7waagyaan ginntiiga-gee-7isan tlada.aaw-gwii ’la 7is rujuu-haal-aa-n-ii. and animal-ART:DEF-too mountain-on 3 go all-CAUS-EVID-PST-TC 'And he told the animals too to go up on a mountain.'

(Enrico 2003: 1342)

Contextually unique referent:

(197) kyaa k’yuw-ce 7inn ’la 7isda-gan. outside door-ART:DEF beside 3 put-PST 'He put it outside by the door.'

(Enrico 2003: 52)
Halkomelem

Referential article

Anaphoric referent:

(198) a. ƛ́á ce’ kʷáx-t [kʷθə smǝɣəθ] wə-mí-əs cə’ be.3 FUT shoot-TR ART:REF.NVIS deer when-come-SUB:3 FUT wə-čəqʷ-θət already-pass.through-self
‘They will be the ones who will shoot the deer when they start coming through.’

b. wə-mí-əs wəłwéł-at-əs [kʷθə sməɣəθ]
when-come-SUB:3 be.chasing-TR-SUB:3 ART:REF.NVIS deer
‘...when they come chasing the deer.’

Contextually unique referent:

(199) łə́q́ ˀə ḥə́q́íq́əłə́s [tə ḱən-sqʷəmə́y]?
habitually Q be.biting ART:REF.VIS your-dog
‘Does your dog bite?’

Specific referent:

(200) s-ni’-ct wə-kʷá-c-əkʷ [tə čičí’qən]
NMLZ-AUX-OUR PFV-see-TR ART:REF.VIS mink
‘Then we saw a mink.’

Nonspecific referent:

(201) nə-s-c-ƛ́íˀ [kʷə páy]
my-NMLZ-do.value ART:REF.REM pie
‘I want some (any kind of) pie.’
Koasati

Anaphoric article

(202) má:mo:si-n st-olá-ci-n iko:si-sáya im-alá-k
be.then-DS INSTR-there-3PL-DS aunt-ART:ANA poss:3-brother.in.law-s
háki-hco:li-sk hisakmápk-on á:t-illí-sáy-on isbakkí
be.drunk-CUSTOM-CONSEQ pull.by.hair-DS.FOC person-dead-ART:ANA-o.FOC head
st-itta-folomóhli-n.
INSTR-LOC-wave.about-DS
'And then, they brought him over there, and the brother-in-law of this aforementioned
aunt, being drunk, grabbing the aforementioned dead man by the hair, was waving his
severed head about it.'  

(Kimball 1985: 345)

Lakota

Definite article

Anaphoric referent:

(203) a. Tok’a wetu ki lejaŋl 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. na [witka ki] nakuŋ tuktekel oyas’iŋ kahuhugapi iš c’iŋpi c’aŋna
and egg ART:DEF also sometimes all break.PL FOC want.PL when
oyas’iŋ aglipi na iyohpiewic’akiciyapi c’aŋna taŋtaŋtunyan glutapi s’a na
all bring.PL and cook.3PL.BEN when excessively eat.POSS.PL ITER and
nakuŋ ziŋtkala ol’ota wic’aopí.
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)

Contextually unique referent:

(204) [Lak’ota ki] ehaŋni Šahiyela na Sapa Wic’asa ob kic’izapí ...
Lakota ART:DEF long.ago Cheyenne and Black Man with fight.PL ...
'In the old days the Lakota used to fight the Cheyenne and Utes ...'

(Ingham 2003: 99)
Exclusive-specific article

(205) [igmu’ wâ] wachi
   cat ART:EXSPEC want.s:1SG.o:3SG
   'I want a (particular) cat.' (Van Valin 1977: 64)

Nonspecific article

(206) [igmu’ wâži] wachi
   cat ART:NSPSEC want.s:1SG.o:3SG
   'I want a (any) cat.' (Van Valin 1977: 64)

Nuuchahnulth (Definite article)

Anaphoric referent:

(207) a. ʔoˑcaḥtaksa ńicwǐsa naŋaqnakšiʔaƛ meʔiƛqacʔisok.
   an.then get.pregnant give.birth.to.baby.now little.boy.poss3
   'Now then she became pregnant and gave birth to a baby, a little boy.'

   b. ńoˑcsme-te ƛaˑḥmał-ʔi
   woman-ART:DEF newborn-ART:DEF baby
   naŋaqak ʔani ḥiḥiŋqəqawat
   that be.bloody.in.fingernails.poss
   čałča baŋqakšiʔol-ʔi
   fingers baby.that.become.poss-ART:DEF
   'Now the woman noticed that the new-born baby that she had given birth to was
   bloody under his finger-nails, so then she began to suspect that it might be he who
   was causing people to die off.'
   (Sapir & Swadesh 1939: 14)

Contextually unique referent:

(208) ʔaˑqac’al ḥuˑcsma=aˑk ḥuˑxwičit
   visible-CAUS.PFV=TEMP woman=poss Woodpecker
   qawaˑs-sac=ʔi
   salmonberry-container.poss-ART:DEF
   'Woodpecker’s wife brought out her salmonberry dish.' (Davidson 2002: 111)
Pipil

Definite article

Anaphoric referent:

(209) a. kunih yaha ki-chiwi-ki se: trampah.
then he 3sg-make-pst art:indef trap
‘Then he made a trap.’

b. tik [ne trampah] weli-k ki-kutamina se: siwa:pil como de dieciocho
in art:def trap can-pst 3sg-catch art:indef girl as of eighteen
years
‘In the trap he could catch a girl of eighteen years.’

(Campbell 1985: 868)

Contextually unique referent:

(210) entonces [ne i-siwa:-w] ki-mik-tih ne chumpipi, ki-chiwi-ki
then art:def 3sg-wife-poss 3sg-die-caus art:def turkey 3sg-do-pret
desplumar, wan ki-chiwi-ki ne komidah wan ki-kwah-ke-t.
pluck and 3sg-do-pst art:def food and 3sg-eat-pst-pl
‘Then, his wife killed the turkey, plugged it, and prepared the food and they ate it.’

(Campbell 1985: 868)

Indefinite article

Specific referent:

in that time come-pl hurricaners from art:indef town of name Chiltiupan
‘In those times hurricaners came from a town named Chiltiupan.’

(Campbell 1985: 867)

Nonspecific referent:

(212) ni-mits-maka-skiya [se: mu-tamal], pero tesu ni-k-piya.
1sg-2-pl-give-would art:indef 2-tortilla but no 1sg-3sg-have
‘I would give you a tortilla, but I don’t have any.’

(Campbell 1985: 123)
Sahaptin

Anaphoric article

(213) ka’aw=tya awkú i-wachá payklá [kw’ink wat’uyológica].
swift=MOD then s:3SG-COP.PST obedient.one ART:ANA oldest.one
‘He was fast, that aforementioned obedient eldest.’

(Jansen 2010: 198)

Tarahumara

Definite article

Anaphoric referent:

(214) a. echarí rocogó nahua-ri biré taa namuti.
at.that night come-PST INDEF small animal
‘On that night a small animal came.’

b. nijé quetasi machi-ri piri niíro.
1SG NEG know-PST what be.IMPF
‘I didn’t know what it was.’

c. echijiti echi rochí sa-ya, echirigá nahua-ri [echi taa therefore ART:DEF fish smell-PTCP in.this.manner come-PST ART:DEF small namuti].
animal
‘Because of the smelling the fish the little animal came.’

(Cohen 1998: 297,311,317)

Contextually unique referent:

(215) Jena’i nirú baquia taraca ra’icha-lú-ami nijeni nama-li [echi nijé here exist three recount-PTCP speak-PASS-ADJ 1SG.NOM hear-PST ART:DEF 1SG
apalócha-la] mapalí nijé pe taa tohuí niíli.
grandfather-spC when 1SG.NOM slightly small boy be.PST
‘Here are three stories that I heard told by my grandparents when I was little.’

(Cohen 1998: 60)
Tepehua

Inclusive-specific article

Anaphoric referent:

(216) a. 7entons nii paastak-lich juu 7ukxtin nii ka-nawii-ya7 juu then COMP think-PFV ART:INSPEC boss COMP IRR-make-FUT ART:INSPEC barda. 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:INSPEC wall 'Well, they built the wall.'

(Kung 2007: 673)

Contextually unique referent:

(217) 7alhch sast’aanta [juu laqachaqan].
go.PFV.ALD PST-sell-INO-PF ART:INSPEC town 'He went selling in the town.'

Tepehua (Kung 2007: 643)

Specific referent:

(218) 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:INSPEC job 'They yelled that they were going to give him a job.'

(Kung 2007: 463)

Presentational article

Specific referent:

(219) Maa chunch nawii-ta [puma-tam kin-tata7] ... EVID thus do-PFV CL:HUMAN-ART:PRES 1POS-old.man ... 'That’s what an old man did ...'

(Kung 2007: 631)

Nonspecific referent:

(220) [tam maqaali7] ka-maa-ch’ixtaq-ni-nch juu tuumiin aantu qoxiyaa ART:PRES rich.person IRR-CAUS-loan-DAT-O:2 ART:INSPEC money NEG good tuumiin palata. money better 'A rich person could loan you money, but it isn’t good money.'

(Kung 2007: 616)
Ulwa

Definite article

Contextually unique referent:

(221) Kurh-pi balaska ya tak-panih; dâ-pam at yam-pangh. scratch-PROX scab ART:DEF peel.off-IMPER.NEG.2 let-OBV.2 at.once heal-IMPER.3
'Don't scratch the scab off; let it heal.' (T. Green 1999: 101)

Establishing referent:

(222) Pukka bara-ka bahangh, yapa ak-pi yak-tikda di-ki luk-dang night dark-ADJ because thus groove-PROX find-PST.1SG thing-POSS:1SG lost-PERF.3SG dai ya]. PST ART:DEF
'As it was a dark night, I groped around like that and found my thing that had been lost.' (T. Green 1999: 139)

Deictic referent:

(223) [Yaka pan-ka ya] yam-ka dayap-o sakpa-h. that stick-3SG ART:DEF good-ADJ lean-SS stand-SS stand-2SG.IMPER
'Stand that stick up at a good angle.' (Koontz-Garboden 2009)

Presentational article

Specific referent:

(224) Baka-ki kuring as bakan da-pak muih wâk balna wâlik child-POSS:1SG canoe ART:PRES buy INCH-OBV.3SG person other PL only yam yam-dai. use-PRS.3PL
'My child bought a canoe and other people are the only ones who ever use it.' (T. Green 1999: 108)

Nonspecific referent:

(225) Lihwan watah yang dai pâlaih, [rupik as] bakan-tatik dai. money have.STAT 1SG PST IRR radio ART:PRES buy-POT.1SG PST
'If I had money, I would buy a radio.' (T. Green 1999: 102)
Ute

Definite article

Anaphoric referent:

(226)  
a. doctor 'ura-puga-vaachi-'u-'uru ...
doctor.s be-rem-bkgr-he-that ...
'He was a doctor ...'
b. maay-pugay-'u-am, doctor-i 'uway ...
find-rem-3sg-3p doctor-o art:def.o ...
'They finally found him, the doctor ...'

Contextually unique referent:

(227)  
[pe'eqwat'uru] na'achichi 'uway po'o-na tii'a-tu.book.s art:def.s girl.gen 3sg.gen write-nom good-nom
'The book that the girl is writing is good.'

Yurok (Definite article)

Anaphoric referent:

(228)  
then long.ago upriver pst exist old.woman
'A long time ago there was an old woman upriver.'
b. Tue' wo oot ho okw' ue-k'ep'ew.
and there pst exist poss:3sg-grandchild
'She had a grandson there.'
c. Tue' o chahchew ho soo megetohkwom' [kue ue-k'ep'ew].
and 3sg be.difficult pst so look.after.3sg art:def poss:3sg-grandson
'She looked after her grandson with difficulty.'

Contextually unique referent:

(229)  
Kues [kue 'ne-leen]?
where art:def poss:1sg-glasses/eyes
'Where are my glasses?'

---

15Garrett (2010) does not provide glosses; the glosses are mine and additionally based on the Yurok dictionary http://corpus.linguistics.berkeley.edu/~yurok/web/search.php from the “Yurok Language Project” (Garrett 2017).


Zoque (Definite article)

Anaphoric referent:

(230) a. ʔən ceʔ-tam-pa bi šuyu.  
1E wash-1/2pl-inc ART:DEF pot  
'We'll wash it (the meat) in the pot.'

b. ʔən pa-tam-pa šuyuʔ-hoʔ [bi šiš].  
1E put-1/2pl-inc pot-loc2 [ART:DEF meat]  
'We put the meat in the pot.'  

(Johnson 2000: 115-116)

Contextually referent:

(231) kwando ʔəy ni-hamin.ʔəy-wə ke gaʔ [bi papan].  
when 3E pse-x.suf-com that dct3 [ART:DEF devil]  
'When he remembered that that was the devil.'  

(Johnson 2000: 94)

B.6 South America

Aguaruna (Anaphoric article)

(232) a. nuni-kã ihuɨa dii-a ma nunu paŋkã  
hu-i akapi-numa utu-kã akapi-na yu-hu-a-kã  
PROX-LOC liver-LOC enter-SEQ.3:ss liver-ACC eat-APPL-IMPFV-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]  
look-IMPFV-NMLZ enough COP:SUB.3:ss ART:ANA-ACC liver-ACC  
yu-hu-a-kã utu-kã puhu-taĩ.  
eat-APPL-IMPFV-SIM.3:ss go.in-SEQ.3:ss live-SUB:DS  
'As they were looking at it, after enough time it had gone in to eat the liver.'  

(Overall 2007: 557-558)
Akuntsú (Definite article)

Anaphoric referent:

(233) ...aramíra nom aramíra aparapia dow. nako ata ãka iki ø-ape dow ...woman no woman, non.Indian ideo man get that.way water relat-path ideo aparapia-t ... non.Indian-art:def ...

‘...woman no, woman, the Non-Indian shot ...(They) caught men, that way, the river’s path, the Non-Indian shot ...’

(Aragon 2014: 163)

Contextually unique referent:

(234) mepit-et jè jèrom tiri.
child.of.woman-art:def dem:prox dem:dist two

‘These are the daughters, there are two of there.’

(Aragon 2014: 163)

Apinayé (Definite article)

Anaphoric referent:

(235) a. ne pəŋ ka twit ka u pa nè go kamə̃ ujwə ñum õrør nè conj after 2.nom pound 2.nom grind compl conj water iness boil ds boil conj twəm apoj pa. fat come.out compl

‘Then you pound it, put it in a pan, bring it to a boil, then the fat comes out.’

water iness boil.3

‘Then you take the fat (with a spoon), pour it in another pan and place it on heat.’

(Cunha de Oliveira 2005: 351-352)

Contextually unique referent:

(236) pa na pa iꞌmə̃ atɛ [pikap ja] nûr prôm ket.
1.emp hls 1 1.dat 2erg earth art:def lie.on.nf want neg

‘It is I who don’t want you to lie on the ground.’

(Cunha de Oliveira 2005: 86)
Arawak

Definite article

Anaphoric referent:

(237) a. Kenda, waboroko lokoda, bolhekotoa to khota da-boran, and road in lying,down ART:DEF.NHUM animal me-before wa-balhosen, bolhekotoa da-bora. our-leader lying,down me-before ‘And, on the road, was lying an animal, Captain, lying before me.’

b. Firobero! tapir ‘A tapir!’

c. Ken ka-kodibion-kada-n, [to firobero]. and ATTR-bird-PERF-it ART:DEF.NHUM tapir ‘And it, the tapir, has a bird.’ (Pet 2011: 217)

Contextually unique referent:

(238) De to Ebesilio ken na-balhosen-ka jahada de [to sikoa] khona. I be Ebesilio and their-leader-PERF here I ART:DEF.NHUM village at ‘My name is Ebesiliio, and I am their leader here in the village.’ (Pet 2011: 217)

Indefinite article

Specific referent:


Nonspecific referent:

(240) B-amyn-ka [aba jolhi] da-myn? 2sg-have-PERF ART:INDEF cigarette 1sg-BEN16 ‘Do you have a cigarette?’ (Pet 2011: 139)

---

16Pet (2011) does not provide any glossed example of a nonspecific referent. This example is only given with a morpheme segmentation and a translation; I added the glossing line according to the other glossed examples in Pet (2011).
Asheninka Perené (Presentational article)

Specific referent:

(241)   ari  i=kaNt-akot-a=ri  mamaro  i=koshi-tzi  [aparoni  kooya-aniki]
AFF  s:3M=say-appl-real=rel  owl  s:3M=kidnap-real  art:pres  woman-dim
yrohatzi  y=atziri-vet-a-ni.
when  s:3M=person-frus-real-aug
'This was the story of the owl which kidnapped a little girl when it was a person long
ago.'  
(Mihas 2010: 217)

Nonspecific referent:

(242)   arika  i=N-pos-av-ak-ak-ia  o  i=N-tsitoNk-ak-e=ro  [aparoni
when  s:3M=irr-hit-recip-caus-pfv-irr  or  a:3m=irr-hit-pfv-irr=o:3nm  art:pres
kooya]  i=N-kamaNt-ah-e=ro  i=niro.
woman  a:3m=irr-inform-compl-irr=o:3nm  poss:3m=mother
'When they fight each other or hit a woman, he [the child] will notify his mother about
it.'  
(Mihas 2010: 302)

Ayoreo

Inclusive-specific article

Anaphoric referent:

(243)   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.’
(Bertinetto 2009: 54)

Contextually unique referent:

(244)   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)
Specific referent:

(245) ¡Cajire to! ¡Arócojna-queđejna! ¡Arócojna-queđejna-i deji ne! look too alligator-different.M.SG alligator-different-ART:NSPEC.M.SG exist.3 there
'Look there! It is an alligator! There is an alligator right there!' (Ciucci 2016: 485)

Nonspecific article

(246) Mu que ore i-plata-rigui cuse.
but NEG 3PL 3-money-ART:NSPEC.M.SG exist
'But they have no money.' (Ciucci 2016: 356)

Baré (Anaphoric article)

(247) tʃiñu yawi mehéša me-wát’uka [da tʃiñu] i-báraka.
dog angry 3PL.want 3PL-beat ART:ANA.PROX dog 3SG.NF-run
'The dog was angry, they wanted to beat the dog, it ran away.' (Aikhenvald 1995: 24)

Baure (Referential article)

Anaphoric referent:

(248) teč worapik teč ses ro=aseroko-wo [to ses].
dem.m come dem.m sun 3SG.M=be.strong-cop ART:REF sun
'The sun came up and it was getting strong.' (Danielsen 2007: 448)

Contextually unique referent:

(249) 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?" ' (Danielsen 2007: 448)

Specific referent:

(250) 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.' (Danielsen 2007: 429)
Nonspecific referent:

(251) πi=ahač=ri kwe-’i [to erapoe’].
2sg=ask=3sg.f exist-EMPH ART:REF plantain
‘Ask her if there is plantain.’

(Danielsen 2007: 393)

Carib (Indefinite article)

Specific referent:

(252) [uwapoto-mpo amu] nirompy-i rapa.
older.person-DEV ART:INDEF die-RECPST again
‘It happened again that an old person died.’

(Courtz 2008: 145)

Nonspecific referent:

(253) [akuri teràà amu-kon] si-upi-ja.
reed.rumped.agouti already ART:INDEF-PL 1-seek-PRS
‘I am going to try and find some agoutis.’

(Courtz 2008: 188)

Cubeo (Anaphoric article)

(254) ibedita di=ki ira=ki abilité-a-di?
however ART:ANA=CL:TREE big=CL:TREE COP.NEG-do-PST-Q
‘But was he not in that tree (you were talking about)?’

(Chacon 2012: 332)

Kashibo-Kakataibo (Presentational article)

Specific referent:

(255) a=x ka ’ikën [achusi matá] ka is!
that=3:NARR be.3 ART:PRES hill.abs NARR look
‘There is a hill, look!’

(Zariquiey Biondi 2011: 250)

Nonspecific referent:

(256) ’è=x Lima=nu kwan-xun kana [achushi casaca] bits-kë ’i-tsin-a-n.
1sg=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.’

(Zariquiey Biondi 2011: 436)
Mamaindé (Referential article)

Anaphoric referent:

(257)  a. na-ʔai-sen-tu  halaus-jeʔ-let-Ø-nãn-wa.
   POSS:3-go-CL:CONTAINER-ART:REF break-EMPH-PST2-S:3-PST-DECL
   ‘His car broke down.’

(257)  b. halaus-hĩʔ,  na-sen-tu  nãinʔ  ãun-let-Ø-nãn-wa.
   break-TEMP.DS POSS:3-CL:CONTAINER-ART:REF just  leave-PST2-S:3-PST-DECL
   ‘So after it broke, he just left it.’
   (Eberhard 2009: 354)

Contextually unique referent:

(258)  häi ta-wäinta-thã-tu  na-latha-Ø-wa.
   POSS:1-group-CL:GROUP-ART:REF COP-S:3-PRS-DECL
   ‘Those right there, they are my group.’
   (Eberhard 2009: 407)

Specific referent:

(259)  sıu-tu  wes-khit-latha-Ø-wa.
   basket-ART:REF make-S:1PL-S:3-PRS-DECL
   ‘We are making baskets.’
   (Eberhard 2009: 389)

Nonspecific referent:

(260)  kajat-tu  jain-sitoh-ta-latha-Ø-wa.
   corn-ART:REF eat-want-O:1-S3-PRS-DECL
   ‘I want to eat corn.’
   (Eberhard 2009: 392)

Mapudungun (Presentational article)

Presentational article

Specific referent:

(261)  [kiñe  korona]  amá  ye-nie-tu-rke-la-y-unga?
   ART:PRES crown  CONF CARRY-PROG-ITER-QUOT-IND-1NS-DU NOST
   ‘Aren’t you aware that we wore a crown?’
   (Smeets 2008: 247)
Nonspecific referent:

(262) iñché uam-ün [kiñe mákina katrü-ke-lu ka ketra-ke-lu].
I need-1SG ART:PRESENT machine cut-HAB-NMLZ and plough-HAB-NMLZ
'I want one cutting and ploughing machine.' (Smeets 2008: 218)

Nhengatu (Presentational article)

Specific referent:

(263) ae u-mendai [yepe Werekena] irũ.
3SG A:3SG-marry ART:PRESENT Warekena COM
'He married a Warekena.' (da Cruz 2011: 147)

Nonspecific referent:

(264) ti=rame tau-konseguiri tau-estudai kua sidade kitĩ tenki tau-riku
NEG=SUB A:3PL-find A:3PL-studio DEM:PROX city ALL OBLIG A:3PL-have
[yepe parte] tau-viveri=rã.
ART:PRESENT space 3PL-live-SUB
'If they do not find a studio in the city, we have to have a place for them to live.' (da Cruz 2011: 497)

Parecis (Presentational article)

Specific referent:

(265) a. hoka eye saude=tyo kafaka ...
COORD this health=TOP yesterday ...
'The health issue now ...

b. maniya-tše heṭati instituto tropico nea-ze [hatya instituição]
side-CL:SMALL in.the.old.days institute Tropico name-NMLZ ART:PRESENT institution
itsoa hoka maiha waiye ha-ita wi=kakoa.
come.in COORD NEG good work-cont 1PL=COM
'First entered an institution called Tropico which did not work well with us.'
(Barros Brandão 2014: 424-425)

---

17 Original Portuguese translation: “Ele casou com uma Warekena.”
18 Original Portuguese translation: “Se não conseguirem estudar na cidade, tem que ter alguma parte para eles viverem.”
Nonspecific referent:

(266) [hatya babera] associação maiha nali aka hoko hatyo hatyo z=aokaka=tyo
    ART:PRES paper association NEG LOC have COORD that that 2PL=say=TOP
    imoti certidão.
    non-Indian certificate

The organization did not have any document, that one which the non-indigenous people
call certificate.'

(Barros Brandão 2014: 427)

Sabanê (Referential article)

Anaphoric referent:

(267) iawa-mi san t-osan.
    cl:bark-ART:REF catch o:1-give
    'Catch it (a ring, mentioned earlier) and give it to me.'

(Antunes de Araujo 2004: 93)

Contextually unique referent:

(268) iawa katataliʔ-mi yeyi-al-i
    cl:bark white.man-ART:REF stay-PRES-ASSERT
    'The white man stays on my boat.'

(Antunes de Araujo 2004: 119)

Specific referent:

(269) Manoel anose-mi t-osan-ntal-i
    Manoel bowl-ART:REF 1.o-give-PST-ASSERT
    'Manoel gave me a bowl.'

(Antunes de Araujo 2004: 176)

Nonspecific referent:

(270) kapune-mata-mi palann-dana
    paca-DIM-ART:REF not.have-PRES-EVID
    'She does not have a small paca.'

(Antunes de Araujo 2004: 92)
Wayuu

Definite article

Anaphoric referent:

(271) a. So’u wanee ka’i, a’waatūshima’a wanee jintüi: “Koou, koou, happen.3SG.PST ART:PRES day shout.PST.3SG ART:PRES boy koou koou jalaina je’raiwa tüü”. come.IMPER see this ‘One day, a boy shouted: “Koou, koou, come to see this”’. 19

b. Chi-jintüi-kai, mayaasü ni’rüın türa, asakitshi: ”¿Jaralü yalaka yala?” m.SG-boy-ART:DEF as.soon.as/when saw asked who be here here ‘Seeing that, the boy asked: “Who is there?” ’ 20

(Álvarez 2017: 92)

Contextually unique referent:

(272) Piiraka, jotusu ma’in areepa-kalü!
look burnt very arepa-ART:DEF.SG
‘Look, the arepa is pretty burnt!’ 21

(Álvarez 2017: 13)

Presentational article

Specific referent:

EXIST.PRS.M ART:PRES donkey-DIM POS:1PL house
‘We have a little donkey at home.’ 22

(Álvarez 2017: 32)

Nonspecific referent:

(274) ¿Nna’aya [wanee süü] ta-püla-jatü pejeru’u pü-nain?
there ART:PRES chinchorro 1SG-for-FUT close 2SG-attentive
Is there a chinchorro for me close to you? 23

(Álvarez 2017: 77)

19Original Spanish translation: “Un día un niño gritó: ‘Koou, koou, vengan a ver esto’.”
20Original Spanish translation: “El chico, al ver eso, preguntó: ‘¿Quién está ahí?’ ”
21Original Spanish translation: “¡Mira, la arepa está muy quemada!”
22Original Spanish translation: “Nosotros tenemos un burrito en la casa.”
23Original Spanish translation: “¿Allí hay algún chinchorro para mí cerca de ti?”
Bibliography

Álvarez, José. 2017. Compendio de la gramática de la lengua wayuu.


Brown, R. McKenna et al. 2006. *¿La Útz awäch? Introduction to Kaqchikel Maya language*. Austin, TX: University of Texas Press.


É. Kiss, Katalin & Orsolya Tánczos. 2017. From possessor agreement to object marking: the grammaticalization path of the Udmurt -(j)ez suffix. Budapest.


McIntosh, Justin Daniel. 2011. *Grammatical sketch of Teotepec Chatino*. Austin, TX: The University of Texas at Austin dissertation.


