Articles across the world’s languages

Laura Becker

Universität Leipzig

Friedrich-Schiller-Universität Jena
25/01/18
Defining articles
  .Articles as comparative concept
  .Their function
  .Their domain
  .Their distribution

Grammar mining. General trends
  .Macroareas
  .Article types
  .Article inflection
  .Articles and other elements in the NP
  .Articles and their lexical sources

Summary
In order to compare articles cross-linguistically, one needs a definition of articles as comparative concept. (Haspelmath 2010)

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 concepts, general formal concepts, and other comparative concepts.” (Haspelmath 2010: 665)
I propose three types of criteria to define articles as comparative concept:

- **function** (referentiality)
- **domain** (noun phrase)
- **distribution** (systematicity)

In addition, we need to distinguish between

- definite articles and adnominal demonstratives
- indefinite articles and the numeral *one*
- nonspecific articles and negative polarity items
The function of articles:

Referentiality
### The referential statuses

<table>
<thead>
<tr>
<th>Status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>definite</strong></td>
<td>Both speaker and hearer can fully identify the referent. e.g. European definite articles</td>
</tr>
<tr>
<td>- deictic</td>
<td>Identifiability by ostension. <em>(demonstratives)</em></td>
</tr>
<tr>
<td>- anaphoric</td>
<td>Identifiability by previous mention. e.g. Wardaman, Lango, Runyankore, Abui, ...</td>
</tr>
<tr>
<td>- bridging</td>
<td>The nominal is identifiable because it is part of the frame of a preceding referent. ??</td>
</tr>
<tr>
<td>- unique</td>
<td>Identifiability by (situational) uniqueness. there seems to be no article marking only this function</td>
</tr>
<tr>
<td><strong>specific</strong></td>
<td>The nominal is linked to a unique referent. Mostly, it is the speaker who is able to identify the referent, while the hearer is not. e.g. Biak, Blackfoot, Arawak, Logba, ...</td>
</tr>
<tr>
<td><strong>nonspecific</strong></td>
<td>The nominal is not linked to any unique referent. Any referent that meets the criteria can be linked to the nominal. Tongan, Biak, Blackfoot, Crow, Q’anjobal</td>
</tr>
<tr>
<td><strong>generic</strong></td>
<td>The nominal refers to all possible referents of a kind. ??</td>
</tr>
<tr>
<td><strong>indefinite</strong></td>
<td>will be understood as non-definite specific and non-specific e.g. European indefinite articles</td>
</tr>
</tbody>
</table>
Examples of definite articles

(i) anaphoric

Lango (Eastern Sudanic, Uganda):

(1) òkélò òbèdò píŋ
Okelo 3SG.sit.PER ground
‘Okelo sat on the ground.’ (Noonan 1992: 242)

... 

(2) òkélò-mérê tê rìŋgò
Okelo-ANA 3SG.THEN run
’Then the aforementioned Okelo ran’ (Noonan 1992: 251)
(ii) bridging

Although bridging is a (recognized as) a relevant referential function (e.g. Clark 1975; Hawkins 1978; Erkü and Gundel 1987; Löbner 1998; Schwarz 2009), there seems to be no particular formal expression (i.e. article) linked to this function in the languages of the world.

(3) Kofi tɔ-ɔ krataa bi. ɔtwerɛfoɔ no fi Kumase.
Kofi buy-pst book SPEC writer DEF come.from Kumasi
‘Kofi bought a book. The author is from Kumasi.’ Akan

(4) Марко ja купи книга-ва. Нејзин-от автор е од Македонија.
Marko it.acc bought.3sg book-def.prox poss.3sg-def author is from Macedonia
‘Marko bought a book; its author is from Macedonia.’ Macedonian

(5) mɔkakɛ àndi kɔtì. ë mɔtilɛli a bɛli ndì mɔtɛ bɛndì.
Mokake buy.3sg book DEF writer.cl1 cl1 be person.am French
‘Mokake bought a book; the author is French.’ Mokpe
(iii) Recognitional use

Recognitional uses involve the definite marking of the referent not based on physical perception or previous mention, but on e.g. shared (or common) knowledge or experience.

This use has rather been observed for demonstratives (e.g. Gundel, Hedberg, and Zacharski 1993: 203; Himmelmann 1996; Himmelmann 1997: 61f; Diessel 1999: 105f), but we find markers which are only used in this function, e.g. in Yapese (Austronesian) and Gooniyandi (Bunuban).

In Gooniyandi, in most cases, the marker is used for referents from shared previous experiences or conversations:

(6) ngooddoo-ngga *ginharndi* goornboo wardgilayi-ngangi moolooddja-yidda
    that-ERG  *DEF* woman she.had.gone.with.us Mulurrja-ALL
    niyaji-ngga barnniginaddi gaddwaroo
    this-ERG  she.returned.me afternoon

‘The woman who went to Mulurrja with us brought me back yesterday.’

(McGregor 1990: 146)
Examples of definite articles

(iii) definite

Kaqchikel (Mayan) has a definite article (ri) that marks anaphoric, bridging, and situationally unique definites.

(7) A: ninwayjo ntäj wäy.
    want.1sg eat tortilla
    ‘I want to eat tortilla.’

B: ri wäy k’o chuchi qaq’
    DEF tortilla EXIST near fire
    ‘The tortilla is in the kitchen.’ (anaphoric)

(8) xk’atz’etkän jun jal tziajchik chila pa tinamit. ri ruwi jmul manutzachik.
    see.pst.1pl indef house old there in village DEF roof totally not.useful
    ‘We saw an old house in the village. The roof was completely destroyed.’ (bridging)

(9) ri ik’ jani nqalaj chpam re aq’a re
    DEF moon much is.visible in dem night dem
    ‘The moon is very bright tonight.’ (unique)
Examples of definite articles

(iv) unique
There seems to be no definite article restricted to unique uses; only definite articles that can also be used for anaphoric and deictic definites seem to be able to mark unique referents.

(10) ó ẓọ́ŋọ́ ́ígbé
2SG NEG.look sun.DEF
‘Don’t look into the sun!’

(11) man tatzu ri q’ij
NEG.IMP look.2SG DEF sun
‘Don’t look into the sun!’

(12) ja ik’ konkaan nch’ech’ona jawra cha’q’a.
DEF moon very shine.bright.3sg this night
‘The moon in very bright tonight.’

(13) me-n-hu ɔsram ?no anumere yi
1SG-NEG-see moon DEF evening this
‘I will not see the moon this evening.’

→ due to iconicity / economy / markedness effects?
→ grammaticalization: extension of definite articles to unique referents as last step?
Definite articles

Across languages, the most common types of articles with definite functions are

1. definite articles (distinct from demonstratives), marking anaphoric, deictic, (situationally) unique referents
2. anaphoric articles (distinct from demonstratives), restricted to anaphoric definite referents

→ e.g. English, Kaqchikel, ...

→ e.g. Lango, Wardaman, ...

→ e.g. Russian, ...

What about other scenarios?

→ Hausa?

→ Indonesian?
Examples of specific articles

Specific (excluding definites)

Most languages with a marker for specific referents have a separate definite (or anaphoric) marker. Therefore, the specific marker cannot be used for definite referents in most cases, as e.g. in Akan (Kwa).

(14) kaa no ye papa bi aa me-hyia no wo Kumase dea car DEF be man SPEC REL 1SG meet 3SG be.at Kumasi POSS ‘The car belongs to a (certain) man I met in Kumasi.’

(15) m’-ayɔŋkofoɔ bi baa-a me-hɔ ɛnora anadwo POSS:1SG-friends SPEC come-PST POSS:1SG-THERE yesterday night ‘Some friends of mine came to my place last night.’

(16) adwuma no aa e-fa-a yeŋ no n-nya m-fasɔo bebre work DEF REL 3SG-take-PST 3PL CD NEG-get profit many ‘The company that hired us does not make enough money.’

(17) me-nya-a anka me-wɔ yere *no/ *bi 1SG-get-PST before 1SG-have wife DEF/ SPEC ‘I wish I had a wife.’
Examples of specific articles

specific (including definites)

This function seems to be really rare, and rather restricted to systems with a specific article, but no separate definite article, as in Blackfoot (Algic):

(18) owa-i
    egg-NONSPEC
    ‘an egg’

(Taylor 1969: 11)

(19) owáá- yi
    egg-SPEC
    ‘an/the egg’

(Taylor 1969: 12)
Are there “anti-specific” articles?

Specific markers are most frequently used (illustrated?) with referents that the speaker can fully identify, while the hearer cannot.

→ Are there markers for referents that are identifiable by the hearer but not by the speaker?

Gamilaraay (Pama-Nyungan, Australia) has a marker that might come close to that:

The “definite marker” -ma combines with interrogative markers to indicate that the speaker does not know the referent, but expects the hearer to do so (Giacon 2014: 173).

(20) minyaaya-\textbf{ma} ngindaay gii-b.aaba-y
where-\textbf{DEF} 2pl get-TOT-PST
Where were you all?  

(Giacon 2014: 174)
Examples for nonspecific articles

Markers that indicate nonspecificity of the referent are rare cross-linguistically.

Q’anjobal (Mayan)

(21) ay mi jun-oq yatut thioxh b’ay yich calle?
EXIST Q a-NSPEC house.of.God at back street
‘Is there a church at the end of the road?’

(22) ay jun yatut thioxh b’ay yich calle.
EXIST a.SPSEC house.of.God at back street
‘There is a church at the end of the road.’

(23) asi’ yul jun-oq tuktuk.
go.IMP in a-NSPEC mototaxi
‘Take a mototaxi (any of the ones in the street).’

(24) tzeb’ach yul jun tuktuk. mayal wawrtej naq tz’umon ch’en.
come.IMP in a.SPEC mototaxi already called cl driver
‘Come taking a mototaxi. I already called the driver.’
So far, I did not find a good candidate for generic articles. The closest are two markers from Crow and Lavukaleve.

**Crow (Siouan)**
Crow has a marker -t that is used for referring to the totality of a class of potential referents, according to the grammar. But it seems to be rather a “totality” marker, focusing on the completeness of the referents.

(25) éehk huchalahúua shoopá-t kuss-chisshíi-wa-hche-k awé
those directions four-**TOT**? GOAL-go.back-1A-CAUS-DECL season
shíishiahe shoopá-t kúh koolá-k
different four-**TOT**? also be.there-DECL
‘I made them go back to those four directions; the four different seasons are there as well.’ (Graczyk 2007: 232)

**Lavukaleve (Solomons East Papuan)**
Lavukaleve has a “demonstrative marker” that is used for unspecific, generic referents (Terrill 2003: 175). Unfortunately, the grammar does not provide good examples to say more about the status of this marker.
Are generic expressions / referents universal at all?

In some languages, there seems to be no good way of expressing general referents (referents of general statements).

→ Is there no way to encode it or is the expression only different from what we expect?

(26) nucha ňechaŋ  mu’uchi  airu  gü
I like/love many dog PL
‘I like many many (all the) dogs.’

(Cikuña)

(27) yū  kainu  jaturiküa  ibebü.
I all transport not.like
‘I do not like cars (in general).’

(Cubeo)
The indefinite article in Awakateko (Mayan)

The indefinite article *jun* in Awakateko is used in indefinite specific and nonspecific contexts.

(28) **at jun tiox stzi’ a’ EXIST INDEF church river.bank water**
‘There is a church at the river banks.’ (indef. specific)

(29) **nawaj jun moy want.1sg INDEF car**
‘I want to have a car.’ (indef. nonspecific)
Examples for referential articles

Rapanui (Austronesian)
In Rapanui, in most contexts except for predicative positions, nouns are accompanied by the marker te: it is used to mark definite, specific, and nonspecific referents. Therefore, it has been labelled referential article in the grammar.

(30) ‘īna he mā’e ha mo u’i iŋa i te kai.
NEG PRED light for see NMLZ ACC REF food
‘There was no light to see the food.’ (Kieviet 2017: 93)

(31) ‘i te no ho iŋa tuai era ’ā te taŋata e tahi te ’īŋoa ko
at REF stay NMLZ ancient DIST IDENT REF man NUM one REF name PROM
tu’uhakararo.
Tu’uhakararo
‘In the old times (there was) a man called Tu’uhakararo.’ (Kieviet 2017: 238)

(32) Ko mate atu ’ana ki te vai mo unu.
PRF die away CONT to REF water for drink
‘I’m dying for water to drink.’ (Kieviet 2017: 237)
Examples for referential articles

Rapanui

However, given that it seems to occur in complementary distribution with the marker *he* used for nouns in predicate position, it is not clear, whether *te* is a referential article or an “argument marker”.

(33)  

He  taŋata tau manu era  
PRED man  DEM bird  DIST  
‘That bird was a human being.’  

(Kieviet 2017: 242)

It could also be a “nominal marker”, since it can occur with verbs in argument positions. Also, lexemes in Rapanui are often ambiguous between being a noun or a verb, which is determined by context.

(34)  a. He  pōrekoreko te  ŋā  pokī  ‘i Tāhai.  
    NTR born:red  REF PL child at Tahai  
    ‘Children were born in Tahai.’  
    (Kieviet 2017: 76)  

    b. Mai te  hora era  ō’oku  e  pokī  nō  ’ana ...  
    from REF time  DIST  POSS.3SG.O  IPFV child just  CONT  
    ‘From the time when I was a child ...’  
    (Kieviet 2017: 76)
Elements with similar but distinct functions:

What is not considered as article
Article or nominal classifier?

**Ughele** (Austronesian, Solomon Islands)

The grammar distinguishes three articles in Ughele: the personal article *e*, the common article *na*, and the focal article *ai* (the latter stays rather unclear in its function).

(35) a. meke naghe e Peni ...
    and say PROP Peni
    ‘and Peni said …’

(35) b. Zioni na viu le-lea-na
    Zioni comm bird RED-good-ATTR:3SG
    ‘Zioni is a very nice bird.’

Instead of viewing this as an article system, I would analyze it as a classifier system with two classes: proper nouns and common nouns, which both need to be marked by classifiers accompanying the noun.
The accusative-partitive alternation in e.g. Finnish can modify the interpretation of the referential status of the noun. An object in the accusative is rather interpreted as definite, while an object in the partitive gets an indefinite reading.

(36)  

a. Silja joi maidon.  
    Silja drank milk.\text{ACC}  
    ‘Silja drank the milk.’

b. Silja joi maitoa.  
    Silja drank milk.\text{PART}  
    ‘Silja drank (some) milk.’

Such secondary functions of markers will not considered here.
Obviative markers

**Ojibwe** (Algic, Canada)

If there are more than one argument in the third person, they are distinguished by proximate vs. obviative marking on the noun. In certain cases, this choice reflects prominence in the discourse.

“Obviation is used to distinguish third persons in clauses, but may also be emphasize a particular character as a point of reference in a narrative or narrative portion” (Todd 1971: 183)

(37)  
\[ \text{wgi:waabmaan \ dash niw \ zhi:ihiiban niibna bbaa-gomnid.} \]
\[ \text{3SG:PRX.OBV.SAW then THOSE.OBV ducks.OBV many OVB.FLOAT/about} \]
\[ \text{‘There he saw many ducks swimming about.’} \]  
(Todd 1971: 183)
In Mangarrayi (Mangarrayi-Maran, Australia), plural marking in narratives is restricted to prominent and already introduced referents.

(38)  

a. malam ∅-yi-ńi-wa ḍanji-nantsawu  
man 3sg-go-pc-suf end-his  
‘People were going to the room.’  
(Merlan 1989: 94)  

b. njabaranwa jarbiñ-garan wara-warguj ∅-jululu wara-warguj  
two young.man-du pick.up-red ABS-belongings pick.up-red  
wur-manbu-ni-wa garan-gara-bayi malam-garan, dar?ma  
3du-run-pc-suf those-two-foc man-du emerge  
‘(The) two men picked up (their) belongings, picked them up. They went along, those two men, came out (into a clearing).’  
(Merlan 1989: 95)
A marker for “the deceased one”

Koasati (Muskogean, USA)

Koasati has a variety of elements (derived from participle markers) used with nominals that are labelled articles in the grammar. One of those elements is ó:to with the lexical meaning “the deceased”.

It is mainly used referring to referents that existed in the past, but do not any longer. Therefore, it is restricted to specific referents.

(39) yalí am-aw-ó:to-k cimpónc-o:to-k
here 1sg:poss-grandfather-EXIST.PST-SUBJ Jim Poncho-EXIST.PST-SUBJ
ónti-to-y
come.sg-IIIPAST-PHR-TERM
‘Here my grandfather Jim Poncho (who is now deceased) came.’

(40) ...waciná nařihilk-ó:to-n ca-sobáy-ko-k
...English language-EXIST.PST-OBJ 1sg.stat-know-3NEG(iiA)-ss
‘...at that time the English language was unknown to me.’

(Kimball 1985: 349)
The domain of articles:

The noun phrase
The position of articles in the noun phrase

The parameters for the position and form of the article in the noun phrase are:

<table>
<thead>
<tr>
<th>free</th>
<th>bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>German <em>die Katze</em></td>
<td>Armenian <em>katu-n</em> ‘the cat’</td>
</tr>
<tr>
<td>segmental</td>
<td>non-segmental (tone)</td>
</tr>
<tr>
<td>Basque <em>emakume-a</em> ‘the woman’</td>
<td>Mokpe <em>mólélí</em> ‘the food’</td>
</tr>
<tr>
<td>preposed</td>
<td>postposed</td>
</tr>
<tr>
<td>Diyari <em>nhinha kupa</em> ‘the child’</td>
<td>Supyire <em>ŋkù-ŋi</em> ‘the chicken’</td>
</tr>
<tr>
<td>Infixed</td>
<td></td>
</tr>
<tr>
<td>Mayangna <em>waki-ni-sa</em> ‘the banana’</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NP as anchor</th>
<th>noun as anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ute <em>kavayi ságharumu</em> ‘<em>uway</em>  ‘the white horse’</td>
<td>Turkish <em>büyü̈k bir oda</em> ‘a large room’</td>
</tr>
<tr>
<td>on one element</td>
<td>repeated</td>
</tr>
<tr>
<td>Zoque <em>yoya bi mas šaša?</em> ‘the fattest pig’</td>
<td>Runyankore <em>o-mushaija o-murungi</em> ‘the good man’</td>
</tr>
</tbody>
</table>
Referential markers outside the noun phrase:

What is not considered as article
Articles must be in the noun phrase

Although some markers indicate the referential status of some nominal, they do not count as articles here, since they themselves do not occur the noun phrase.

Macedonian object doubling

(41) Не можам да ja најдам вода-та.
NEG can.1SG comp 3SG.F.ACC find.1SG water-DEF
‘I can’t find the water.’

Agreement markers in Nyaturu, Bantu

(42) a. n-a-onaa mwalimu
sm:1SG-pst1-see cl1.teacher
‘I saw a teacher.’

b. n-a-məs-onaa mwalimu
sm:1SG-pst1-om:cl1-see cl1.teacher
‘I saw the teacher.’

(Hualde 1989: 182)
Articles must occur with nouns

Other markers also indicate referentiality, occur in the nominal domain, but not (primarily) with nouns. Such markers are not counted as articles.

Possessive-based definite marker in Tarahumara
Tarahumara (Uto-Aztecan, Mexico) has a marker that indicates referentiality, but mostly modifies nouns in genitive and possessive constructions. Therefore, they are not counted as articles here:

(43) a. nijé ’huéña-ra
   1SG parent-REF
   ‘my parents’

   (Cohen 1998: 130)

b. echi Antonio huiye-ra
   DEF Antonio mother-REF
   ‘Antonio’s mother’

   (Cohen 1998: 137)

c. biré ripurá cusí-rá
   one axe pole-REF
   ‘an axe handle’

   (Cohen 1998: 120)
Articles must occur with nouns

Indefinite marker for pronouns in Gunin
The indefinite marker in Gunin (Worrowran, Australia) is primarily used with pronouns and not with nouns.

(44) bari ngamirri, nginda-ngurru
come it.will.take.us what-INDEF
‘Something is going to happen to us.’ (McGregor 1993: 42)

Germanic adjective forms

(45) der schwarz-e Hund vs. ein schwarz-er Hund ‘a / the black dog’
The distribution of articles:

Obligatoriness vs. systematicity
Do articles have to be obligatory?

Often, to decide whether a marker is an article or not, obligatoriness has been used as criterion.

The definite article in Hausa
Hausa has a definite article, -n that originated from an anaphoric article, but is more and more used in other definite contexts as well.

(46) yaarò-n dà ya tàfi
     boy-DEF REL 3SG.M.PERF.REL leave
     ‘the boy who left’

Hausa (Newman 2000: 145)

However, a noun with no definite article is not necessarily interpreted as indefinite:

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

(Newmann 2000:143)

Therefore, the definiteness marker has been argued not to be obligatory and hence no “article”, as it does not always occur with nouns that receive a definite interpretation (Newman 2000: 143, Lyons 1999: 52, Zimmermann 2008: 419).
Articles need to occur systematically

Can such a marker be a definite article, if it is not obligatory in definite contexts?

→ YES!

Instead of obligatoriness, I would rather use the notion of systematicity.

If the marker is not used with nouns that receive a definite interpretation in certain contexts that can be “predicted” by rules, it can still count as article.
Articles need to occur systematically

. Restrictions due to morphosyntactic properties of the nominal

(48)  
  a. Ich sehe **eine** Katze auf dem **Dach.**  
      I see **ART:INDEF** cat.sg on **ART.DEF** roof  
      ‘I see a cat on the roof.’  
  b. Ich sehe **Katzen** auf dem **Dach.**  
      I see **cat.PL** on **ART.DEF** roof  
      ‘I see cats on the roof.’

. Restrictions due to phonological properties of the nominal

(49)  
  a. **è nga** ‘the sand’  
  b. **∅ ɪkpa** ‘(the) salt’

. Restrictions due to semantic properties of the nominal

(50)  
  (*The) John gave (*the) Mary the book.
Articles need to occur systematically

. Restrictions due to the syntactic context of the nominal (e.g. with adpositions; cf. Lyons 1999: 51)

(51) yɛn-kɔ dwa-m #no!
1PL-go market-into DEF
‘Let’s go to the market!’

(Kakan)

(52) jo’ pa #ri k’ayb’al
   go.HORT to DEF market
   ‘Let’s go to the market!’

(Kaqchikel)

(53) jo’ pa #ja k’eb’äl
   go.HORT to DEF market
   ‘Let’s go to the market!’

(Tz’utujil)
Articles need to occur systematically

Restrictions due to pragmatic or extralinguistic factors?
E.g. Crow indefinites

(54) a. dakáak-kaata-m húu-laa híi-k
bird-DIM-ART.SPEC come-ss reach-DECL
‘a bird came, it reached him’
(Graczyk 2007: 228)
b. axée baláyyiikaashe dúup-eem alúutkaashe áppaa
father bow two-ART:NONSPEC arrow with
día-a-wa-ku-hee?
make-CONT-1B-give-AFF-INTERR
‘father, will you please make me two bows as well as (some) arrows?’
(Graczyk 2007: 230)

(55) a. bachee-lák baa-aash-dée-k
man-ART.INDEF INDEF-hunt-go-DECL
‘a man went hunting’
(Graczyk 2007: 230)
b. éehk bal-héelee-n iisashpít-dak baappeé-k b-aliat-bee-m
that wood-among-LOC rabbit-ART.INDEF 1A.kill-DECL 1A-think-1A-DS
isáa-kaashi-k
large-AUGM-DECL
‘there in the woods I thought I killed a rabbit, but to my surprise, it was
something very large.’
(Graczyk 2007: 231)
Articles need to occur systematically

- Restrictions due to pragmatic or extralinguistic factors?

  e.g. anaphoric articles in Konso (Afro-Asiatic, Ethiopia)

  In Konso (Afro-Asiatic, Ethiopia), the definite article can be omitted in certain anaphoric contexts (in general, it is used in anaphoric contexts, though).

  (56)  kaasa-sit-n=in  karmaa-si?  ?iʃ-ay
       gun-DEF-INSTR=1 lion-DEF  kill-PF.3.M
     ‘I killed the lion with the gun.’ (Orkaydo 2013: 97)

  In a story about a lion, mentioned before example (57), the lion, identifiable at that point, does not need to be marked as such by the definite article:

  (57)  karmaa ka  çapaleeta-asi?  çaraa kaassuma=i kaassaɗ-ay
        lion  and monkey-DEM on  question=3  ask-PF.3.M
     ‘And the lion asked this monkey a question.’ (Orkaydo 2013: 97)
Grammar mining. General trends
The sample

So far, I considered

- 58 languages (divided into the 6 Macroareas used in WALS)
- 85 articles
  - Africa (14/20)
  - Eurasia (11/17)
  - Papunesia (9/16)
  - Australia (7/7)
  - North America (9/15)
  - South America (8/10)
Macroareas
Macroareas

article type

<table>
<thead>
<tr>
<th>Macroarea</th>
<th>ana</th>
<th>def</th>
<th>indef</th>
<th>nonspec</th>
<th>ref</th>
<th>spec</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eurasia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>papunesia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>australia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n.america</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s.america</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Grammar mining. General trends
Macroareas

position of the article

Grammar mining. General trends
Macroareas

Article inflection

<table>
<thead>
<tr>
<th>Macroarea</th>
<th>count</th>
<th>art.infl</th>
</tr>
</thead>
<tbody>
<tr>
<td>africa</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>eurasia</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>papunesia</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>australia</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>n.america</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>s.america</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Macroareas. Interim results

- **Article type**
  We find many indefinite articles in Europe, many specific articles in Africa, and only definite/anaphoric articles in Australia.

- **Article position**
  Articles in Africa follow other elements in the noun phrase and tend to be postposed.

- **Article inflection** (marking of number, gender, case)
  There seems to be no areal tendency.
Article types
From different formal properties of articles, only their compatibility with plural nouns is a relevant factor to distinguish between different articles types.

Definite articles generally are compatible with plural nouns, while indefinite articles are significantly less compatible with them.
Article types can be clustered based on the formal similarities of the articles.

- The articles cluster according to their semantic values (basically: one def and one indef cluster).

- However, this should be due to the compatibility with plural nouns.

- Taking out compatibility with plural nouns, definite and indefinite articles form a cluster for being less restrictive in their semantics.

- This is manifested in their broader syntactic occurrence (in predicate position, with generic nouns).

- Anaphoric, specific, and nonspecific articles are more restrictive, probably less grammaticalized, and therefore also occur in syntactically more restricted contexts.
Article inflection
The form of the article does not influence whether it marks number, gender, and/or case of the noun.
Neither does the position of the article does not influence whether it marks number, gender, and/or case of the noun.
The inflectional pattern of articles seems to depend on the inflectional morphology of the noun itself.

<table>
<thead>
<tr>
<th>as category</th>
<th>number</th>
<th>gender</th>
<th>case</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUN+ART</td>
<td>22</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>NOUN</td>
<td>44</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>ART</td>
<td>9</td>
<td>18</td>
<td>1</td>
</tr>
</tbody>
</table>

- In general, multiple exponence within the noun phrase is avoided.
- **NUMBER** is mostly marked on the noun only, but seems to show less pressure to be marked only on one element.
- **GENDER** is mostly marked on the article only, but can also be expressed on the article alone (West African noun classes).
- **CASE** is mostly marked on the noun (marked on both noun and article in Australian languages).
Articles and other elements in the NP
Generally, definite articles tend to not co-occur with demonstratives, and they fill the same syntactic slot.

However, if they co-occur, this is correlated with them occurring on different sides of the noun.

This suggests that definite articles and they co-occurrence is rather restricted syntactically, and not semantically.
Definite articles with possessives

With possessives, there is no such correlation as with definite articles and demonstratives.

This suggests that definite articles and demonstratives are somewhat more similar (syntactically?) than definite articles and possessives.
Grammar mining. General trends

Articles and other elements in the NP
In general, articles show a similar behaviour as demonstratives and articles with respect to agreement with the noun.
Looking at the marking of number, gender, (and case) on articles compared to attributive demonstratives, adjectives, numerals, and quantifiers, the following tendencies can be observed:

- If the adjective shows agreement, the article also agrees with the noun.
- If the numeral shows agreement, the adjective also tends to agree with the noun.
- Quantifiers seem to agree with the head noun more often than numerals, but less often than adjectives.

A (very) tentative hierarchy for agreement in the NP could look like:

(58) \[ \text{ART} > \text{ADJ} > (\text{QUANT}) > \text{NUM} \]
Articles and their lexical sources
Lexical sources for articles

What other sources are there for articles than the “obvious ones”?

- demonstrative > definite article
- numeral *one* > indefinite article

- personal pronouns
- possessive pronouns
- copulas, stative predicates
- participles
- partitives
(i) definite articles from pronouns in Diyari (Pama-Nyungan)

(59)  
   a. thangkuthangkuparna ngathu nhinha yakalka-yi
       morning-LOC 1SG.ERG 3SG.NF.ACC ask-PRES
       ‘I will ask him in the morning.’  
   (Austin 2013: 91)

   b. thana nhinha nganthi yingki-mali-yi
      3PL.NOM DEF.ACC MEAT.ACC give-RECIP.PRES
      ‘They give each other the meat.’  
   (Austin 2013: 82)

(ii) definite articles from possessive pronouns in Amharic

(60)  
   tämari-w ‘student-POSS / student-DEF’  
   (Appleyard 1995: 62)
Lexical sources for articles

(iii) definite articles from copulas

Arawak (Arawakan, Brazil)

(61) a. lira wadili to da-reti
    that man is my-husband
    ‘That man is my husband.’

b. to kodibio balyta ada dyna diako
   DEF bird sit tree arm on
   ‘The bird sat oon a branch.’

Gaahmg (Eastern Sudanic, Sudan)

(62) a. tɔ̄gg sèggār-g=à
    COW-PL strong-PL=COP.PL
    ‘Cows are strong.’

b. ɔ̄ ɛ̄ mūn nāán bāārg=à ṇáō-àn ṇā-lg nà
   and with time that Baggara=DEF.PL serach.for-CONT.P girl-PL REL
   ōn-g-ì
   young-PL=RDM
   ‘At that time the Baggara were kidnapping young girls.’

(Pet 2011: 43)

(Pet 2011: 44)

(Stirtz 2011: 160)

(Stirtz 2011: 135)
Lexical sources for articles

(iv) anaphoric markers from participles
Koasati (Muskogean)

   orange eat-1SS-PTCP-FOC be.good-ADV-PST
   ‘The orange I am eating is very good.’ (Kimball 1985: 245)

   b. ...ikoːsi-ːsáya im-alá-k ...
      ...aunt-ANA 3POSS-brother.in.law-SUBJ
      ‘...the brother-in-law of this aforementioned aunt ...’
      (Kimball 1985: 345)

(v) indefinite plural articles from partitives
Ch’ol (Mayan)

(64) a. li ńox-ob-tyak
   DEF elder-PL3-PART
   ‘some of the elders’
   (Vázquez Álvarez 2011: 123)

   b. x-ja’a-ts’i’-tyak
      NCL-water-dog-INDEF
      ‘some nutrias’
      (Vázquet Álvarez 2011: 243)
The marking of number, gender, and case on articles seems to be largely determined by the inflection of its source lexeme, i.e. mostly number in this case. Articles do not seem to develop agreement morphology. They retain the morphology of their lexical source, or lose agreement marking.
Also for definite articles from demonstratives, we see that articles retain or lose agreement morphology, but do not develop new marking.

<table>
<thead>
<tr>
<th>language</th>
<th>ART type</th>
<th>ART inflection</th>
<th>DEM inflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supyire</td>
<td>def</td>
<td>num;gen</td>
<td>num;gen</td>
</tr>
<tr>
<td>Konso</td>
<td>def</td>
<td>num;gen</td>
<td>num;gen</td>
</tr>
<tr>
<td>Sheko</td>
<td>def</td>
<td>gen</td>
<td>(num);gen</td>
</tr>
<tr>
<td>Tafi</td>
<td>def</td>
<td>/</td>
<td>num;gen</td>
</tr>
<tr>
<td>Hungarian</td>
<td>def</td>
<td>/</td>
<td>num;case</td>
</tr>
<tr>
<td>Armenian</td>
<td>def</td>
<td>/</td>
<td>num;case</td>
</tr>
<tr>
<td>Maori</td>
<td>def</td>
<td>num</td>
<td>num</td>
</tr>
<tr>
<td>Alamblak</td>
<td>def</td>
<td>/</td>
<td>num</td>
</tr>
<tr>
<td>Abui</td>
<td>ana</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Ch’ol</td>
<td>def</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>
Inflection of articles originating from the numeral

Again, we see the same tendency for indefinite articles originating from the numeral *one*.

An exception is the article in Konso, where the indefinite article adapts to the gender marking on other elements in the noun phrase, e.g. the definite article.

<table>
<thead>
<tr>
<th>language</th>
<th>ART type</th>
<th>ART inflection</th>
<th>NUM inflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logba</td>
<td>spec</td>
<td>num;gen</td>
<td>num;gen</td>
</tr>
<tr>
<td>Konso</td>
<td>spec</td>
<td>gen</td>
<td>/</td>
</tr>
<tr>
<td>Tafi</td>
<td>spec</td>
<td>num;gen</td>
<td>num;gen</td>
</tr>
<tr>
<td>Hungarian</td>
<td>indef</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Albanian</td>
<td>indef</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Armenian</td>
<td>spec</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>German</td>
<td>indef</td>
<td>gen;case</td>
<td>gen;case</td>
</tr>
</tbody>
</table>
Inflection of articles: economy vs. diachrony

- the inflection of the article depends on the inflection of the noun
- multiple exponence is avoided

- the inflection of the article depends on the inflection of its lexical source
- the article can lose morphology, but does not develop new paradigms
Summary

Article types

- The type of article and most other formal properties is not majorly biased by the geographical area.
- The only significant factor to distinguish between article types is their compatibility with plurals.

→ Despite their differences, it we see that articles indeed behave like elements of one category cross-linguistically.

Inflection

- The inflection of articles depends on morphological properties of its source.
- Also, the inflection of articles depends on the morphological marking on the noun: number is quite frequently expressed on both, gender has a strong preference for being marked on the article only, while case is almost always marked on the noun only.

Other elements in the NP

- The co-occurrence of definite articles and demonstratives seems to follow syntactic restrictions, if they occur in the same position, they tend not to co-occur.
- With possessive pronouns, we do not find such a correlation.

→ Demonstratives and articles seem to be more alike than possessives.

- There is a tendency for articles to be inflected more frequently than adjectives, quantifiers, and numerals.
Appendix
Definite article or demonstrative?

It is difficult to draw a well-motivated consistent line between demonstratives and definite articles.

The contexts of (spatial) deixis, anaphora, and uniqueness of the referent can be viewed as a continuum from demonstrative to article, from less grammaticalized to more grammaticalized, etc.

In this vein, markers that code anaphoric reference are not always viewed as articles, but demonstratives:

“But this means that it may be unclear whether a determiner specialized in anaphoric use is an article or a demonstrative” Lyons (1999: 54)
I define the cut-off point as follows:

<table>
<thead>
<tr>
<th>sensory perception</th>
<th>anaphoricity</th>
<th>uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM</td>
<td>DEM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ART</td>
<td>ART</td>
</tr>
</tbody>
</table>
Indefinite article vs. the numeral *one*

The numeral *one* is the most frequent source for indefinite articles. In addition, even though a language might not have an indefinite article, the numeral can indicate non-identifiability of a referent that would be interpreted as salient, prominent, and thus identifiable otherwise.

→ How can we tear apart those two element and uses? When can we count such an element as article?

1. if it has different formal properties than the numeral *one*:
   - its morphological form is different
   - its position in the noun phrase is different
   - it has a different stress / prosody pattern

2. if it is compatible with the plural (while the numeral is not)

3. if none of the conditions listed above apply, it can still be defined as article based on its function and distribution:
   - it occurs frequently in context such as introducing new relevant, but not identifiable entities into the discourse
   - it occurs with different noun types (animate, inanimate, abstract, etc.)

4. if non of these conditions apply, I do not consider the marker to be an article.
In some cases, it is not trivial to decide whether we deal with one article, that has a separate semantic category with different values it can take, or if we should rather consider those separate articles, with an additional function each.

### Macedonian definite article(s)

<table>
<thead>
<tr>
<th>“distance”</th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>DIST.NEUTR</td>
<td>ot</td>
<td>ta</td>
</tr>
<tr>
<td>PROX</td>
<td>ob</td>
<td>ba</td>
</tr>
<tr>
<td>DIST</td>
<td>oh</td>
<td>ha</td>
</tr>
</tbody>
</table>

### Abui anaphoric article(s)

<table>
<thead>
<tr>
<th>“distance”</th>
<th>speaker-oriented</th>
<th>hearer-oriented</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROX</td>
<td>do</td>
<td>to</td>
</tr>
<tr>
<td>MED</td>
<td>o</td>
<td>yo</td>
</tr>
</tbody>
</table>
Since these properties do not impose major changes on the referential function of the articles, I analysed them as one article with its own category, whose values are marked by the different exponents.
“Rare” properties of articles

In Abui, the anaphoric article is marked for speaker or hearer anaphoricity.

**Abui** (Timor-Alor-Pantar)

(65)  

a. pelang o  tut tah-a  it-i canoe **ANA.1** shore put.on.cpl-dur lie.on-pfv  
‘that canoe (I talked about before) lies on the shore’

b. na kul we ne-sura  **to**  he-bilen-r-i=te  
1sg must leave 1sg.al-book **ANA.2** 3ii.loc-colour-reach-pfv=incp.c  
‘I must go and photocopy my book (that you just talked about) first’

(Kratochvıl 2007: 115)
“Rare” properties of articles

3sg pronouns as articles in Papuan Malay

Papuan Malay (Austro-Asiatic, Papua New Guinea) uses 3sg pronouns adnominally to mark the definiteness of the noun.

(66) a. de bilang, ko tidor apa?
3SG say 2SG sleep what
‘He said, “why are you sleeping”? ’

b. baru nene de mulay tanya saya
and.then grandmother DEF start ask 1SG
‘[...right then (I) met my grandmother, grandmother and then (my)
older brother, aunt’s child.
And then grandmother started asking me ...’

(Kluge 2017: 292)  

(Kluge 2017: 355)
Articles inflecting for person in Biak
In Biak (Austronesian), articles are marked for person and can be combined with first and second person as well.

(67)  mananwir  an-mko-ya
village.chief DEF-2PL-SPEC
‘you village chiefs’

(68)  a.  wudu-wi Ø-ba-ndi  wuja-yi
  little-DEF 3SG-burn-PST fire-INSTR
  ‘When he was little he got burned.’

       (Merlan 1994: 93)

       b.  bulju-wi  Ø-yanggi  yiwarna  gandawag
  long.ago-DEF 3SG-go.PST other.ABS moon.ABS
  ‘He went long ago, last month.’ (Merlan 1994: 93)
Articles with pronouns in Rapanui
In Rapanui, the definite article marking proper nouns (if it can be considered article) also co-occurs with personal pronouns.

(69) Ka rima ta’u a tu’a he mana’u haka’ou a ia ki a Roke’aua CNTG five year by back NTR think again DEF 3SG to DEF Roke’aua ararua ko Makita. the.two PROM Makita ‘Five years later he thought again of Roke’aua and Makita.’

(Kieviet 2017: 129)

Elevated genre articles in Crow
In Crow (Siouan, USA), definite and indefinite articles are replaced by a single marker in an elevated narrative genre, neutralizing the difference in referentiality otherwise indicated (Graczyk 2007: 233).
If articles, as well as case markers are realized as suffixes on the noun, we find unexpected (?) variation:

**NOUN-ART-CASE** (Dime, Maricopa, Basque)

(70) ʔaté guur-af-is-im deis-i-t
1SG.NOM crocodile-PL-ART:DEF-ACC kill-PF-1
‘I killed the crocodiles.’

(71) iipaa-ny-sh hrnii-k
man-ART:DEF-NOM tall-ASP
‘The man is tall.’

**NOUN-CASE-ART** (Armenian, Wardaman)

(72) dasaxos-ě usanol-i-n tvec’ girk’-ē
‘The lecturer gave the book to the student.’