Efficient design
Occam’s razor leads to the efficient design hypothesis
One generally overlooked point: efficient design should make language easily learnable
But what do we know about efficient design?
Why not binary computer language systems or finite state devices?
What (the most) effective design is is not given a priori. Also, effective design for what?

Second position clitics (see Bošković 2016a)
Second position clitics in SC (the only possible placement of clitics in these examples)¹

(1)  a. Mi/ Zašto smo mu je predstavili juče.
    we/why are him.dat her.acc introduced yesterday
    ‘We introduced her to him yesterday./Why did we introduce her to him yesterday?’

   b. Ona tvrdi da smo mu je mi predstavili juče.
    she claims that are him.dat her.acc we introduced yesterday
    ‘She claims that we introduced her to him yesterday.’

   d. Predstavili smo mu je juče.
    introduced are him.dat her.acc yesterday
    ‘We introduced her to him yesterday.’

What is responsible for the availability of second position clitic systems crosslinguistically?

The second position clitic generalization

(2) Second position clitic systems are found only in languages without articles.

The difference between English (3) and Serbo-Croatian (SC) (4) is standardly assumed to be PF-based,
the only difference between English and SC being that D is phonologically null in SC.

(3) The stone broke the window.
(4) Kamen razbi prozor.
    stone broke window       (SC)

Bošković (2008a, 2012): There is a fundamental syntactic difference in the traditional Noun Phrase (TNP)
of English and languages like SC that lack articles which can be captured if DP is not even present in the
2015, Bošković & Hsieh 2013, 2015,a.o for no-DP analyses of individual article-less languages).

Definite articles (Slovenian, see Bošković 2009)

¹Note that true second-position clitics are not simply enclitics, see Bošković (2001).
The article is unique (a distinct form, occurs only once per TNP).
It roughly has the meaning of an iota-operator, yielding an element of type e (see Bošković and Hsieh 2015). Given Chierchia’s (1998) proposal that type shift from type <e,t> to type e is possible in a language only in the absence of a definite article, which means bare NPs can have definite interpretation only in languages without definite articles (i.e. NP languages) what is considered to be a definite article must be present for definite interpretation in a DP language/language with a definite article (and yield such interpretation). Notice that bare NPs “cats” and “window” can have e-type interpretation in SC (4), which in English requires the presence of the (3).

A preliminary list of second position clitic languages, to be expanded: a number of Slavic languages (SC, Czech, Slovak, Slovenian, Hucul Ukrainian, and Sorbian), Latin, Ancient Greek, Pashto, Tagalog\(^2\), Ngiyambaa, Warlpiri, Ossetic, Northern Talysh, Southern Tati, Comanche, Hittite, and Sanskrit.\(^3\)

Restating (2):\(^4\)

(5) Second position clitic systems are found only in NP languages.

**Slavic:** while a number of Slavic languages have second-position clitic systems, Bulgarian and Macedonian, the only Slavic languages with articles, are glaring exceptions.

**Romance:** Latin had second-position clitics, while Modern Romance languages lack them.\(^5\)

**History of Greek:** Taylor (1990) shows that 90% of enclitics in the Homeric period, when Greek did not have articles, were in the second position; this simple second position cliticization system broke down in the later stages (i.e. article stages), like Koine Greek.

**Ossetic:** a Northeast Iranian language with two distinct main dialects (they are mutually barely

\(^2\)Tagalog –ang is not a definite article (Kroeger 1993, Rackowski 2002, Aldridge 2004, Rackowski and Richards 2005, Wurmbrand 2013). It is also not obligatory for definite interpretation; note the ambiguity of the object in (i).

(i) Sino ang b-um-ili ng damit?
   who ANG Nom.asp.-buy CS dress
   ‘Who is the one who bought a/the dress?’ (Nakamura 1996:56)

\(^3\)Regarding less known cases, for Comanche see Steele (1977), Charney (1993), McDaniels (2008); for Ossetic, Abaev (1964), Erscheler (2012); for Northern Talysh, Cysouw (2003, 2005), Paul (2011); for Southern Tati, Yar-Shater (1969). The classification of Southern Tati is a bit tentative since the discussion in Yar-Shater (1969) is not comprehensive enough and glosses are not given. Some relevant examples from different dialects of Southern Tati, with the glosses provided:

(i) azir-öm äš bepat
   yesterday-1sg äš cooked
   ‘Yesterday I cooked äš’

(ii) ay-im bind
   him-1sg saw
   ‘I saw him.’

(iii) Em amberâzz-öm sanduq-u andâs.
   this dress-1sg trunk-in found
   ‘I found this dress in the trunk’ (Châli)

(iv) dō berâ-š da.
   two brother-3sg had
   ‘he had two brothers’

(v) deraxt-i bâši
   tree-2sg fell
   ‘You(sg.) felled the tree.’ (Xoznini)

(vi) cemen-i orosiehâ bad beduta me.GEN-2sg shoes poorly sewed
   ‘You have sewed my shoes poorly’ (Eshtehardi)

\(^4\)It’s possible (2)/(5) will turn out to be strong tendencies, which would still call for an explanation. The deduction of (2)/(5) proposed in Bošković (2016a) actually leaves room for such a scenario (see also Bošković 2016a on Chamorro).

\(^5\)Old Spanish was not an exception, see Wanner (2001).
intelligible, see Thordarson 1989), Iron (or East Ossetic or Tagauric) and Digor (or West Ossetic).
Abaev (1964): the two differ with respect to articles; Digor has a definite article but Iron does not.
Erscheler (2012): Iron is a strict second-position clitic language, Digor is not.
Illustration: both Iron and Digor are multiple wh-fronting languages, where non-D-linked wh-phrases cluster together in front of the verb. Importantly, clitics intervene even between fronted wh-phrases in Iron (but not in Digor) due to the second position requirement.

(6) či=ma=šən  sə  żonə  asə  fišivad-ən?
who=also=DAT.3PL  what  know.PRS.3SG  this  youth-DAT
‘Who knows what about them, about this youth?’  (Ajlarty 2002:13, apud Erchsler 2012:678)

Pama-Nyungan languages
The following languages have second position clitic systems (either strict second position clitic systems, or second position clitic systems with some exceptions; the sources are Cysouw 1993, Dench 1998, Meakins & Nordlinger 2013, Mushin 2005a,b; 2006, McConvell 1996, Alpher 1991)
Yingkarta, Wajarri, Ngiyamba, Warlpiri, Warumungu, Bilinarra, Warnman, Nhanda (only subject clitics), Pitjantjatjara, Yir-Yoront, Gurindji, Djaru, Ngarinyman, Mudburra (undergoing a change), Wembawemba, Wergaia, Madimadi, Wathawurrun, and Woiwurru.
Yukulta, Garrwa, and Wambaya, non-Pama-Nyungan Australian languages, also have 2\textsuperscript{nd} position clitics.


Some illustrations where bare NPs receive an e-type reading.

(7) mayu  njinanja  parnangka
child-ABS  sit-PST  ground-LOC
‘The child sat on the ground’  (Wajarri, Douglas 1981:230)

(8) Alaji  buguwa-nguji  darranggu-nguji.
boy:I(NOM)  stick:IV:Abs-PROP:I(NOM)  stick-PROP:I(NOM)
‘The boy has a big stick.’  (Wambaya, Nordlinger 1993:138)

(9) birrkaliiba=ngayu  waliyi-nyi,  winjawa  näyi  nganyi  wulkanja  waliyi-yudi
hungry=1sg  meat-DAT,  where  here  your  father  meat-PROP
‘I’m hungry for meat. Where’s your father with the meat?’  (Garrwa, Mushin 2005b:263)

(10) rtangka-ya=ka-rrī  ngawu  pala-tha
man-ERG=TR=PRES(R)  dog(ABS)  hit-IND
‘The man is hitting the dog. [ACTIVE]’  (Yukulta, Keen 1983:206)

(11) nyaru-lu-ngu  yawarda  nha-‘i
woman-ERG  kangaroo.ABS  see-PAST
‘The woman saw the kangaroo.’  (Nhanda, Blevins 2001:48)

(12) Billy-ⅷ tjiṭji  nya-ngu
Billy-erg  child  see-past
‘Billy saw the child.’  (Pitjantjatjara, Aissen 2003:452)

Only one case where a language from the above group was claimed to have a definite article: WALS classifies Yingkarta as a language with a definite affix (–ja), based on Dench (1998)
This classification is incorrect (see also Austin 1995, 2006, who treats Mantharta –*thu*, which Dench 1998 says is a cognate of –*ja*, as a topic marker).

–*ja* is not obligatory for definite interpretation, as shown by (13), where –*ja* is not present. Furthermore, (14)-(19) show that –*ja* can be used with pronouns (14)-(15), adverbs (16), adverbial wh-phrases (17), and verbs (18)-(19) which indicates that it is not a definite article.

(13) Thuthu-ngku jarti-lanyi mantu.
    dog-erg        eat-pres     meat
    ‘The dog is eating the meat.’

    (Dench 1998:22)

(14)     Kurra-rtu   mangu   nyina-angkulpa  nganh u-ja.
    not-1plS  good  sit-IMPF  1plNOM-DEF
    ‘We’re not good (well, happy), staying here.’

    (Dench 1998:40)

(15)     Thuthi-lkarangu milyura,  wirntirina-warangu,  pika-piya-warangu nyinta-ja.
    tread.on-APPR snake bite-APPR sick-INCH-APPR 2sgNOM-DEF
    ‘You might tread on a snake, (it) might bite (you), you would get sick.’

    (Dench 1998:76)

(16)     Wanthapara-rtu nyina-angku,  mangu-ja?
    how-1plS   sit-IMPF  good-DEF
    ‘How will we be (after this wind stops), good?’

    (Dench 1998:44)

    when-DEF   later-DEF       not      now-DEF
    ‘When (are you going)? Later, not now.’

    (Dench 1998:70)

(18)     Ngurlupiya-nyi-ja maru-ngka yana-wara.
    fear-PRES-DEF night-LOC go-PURP
    ‘(They’re) frightened to go at night.’

    (JD) (Dench 1998:30)

    that-DEF  swim-PRES-DEF (river)-LOC run-RELds   child
    ‘The children are swimming in the river which (while it) is flowing.’

    (Dench 1998:72)

**Uto-Aztecan languages** (with second position clitics)

(20) **Northern Uto-Aztecan languages**

**Numic languages**


**Takic languages**

Cupeño (second position subject clitics, second position aux clitics, Steele 1977, Hill 2005), Luiseño (second position subject clitics, aux clitics, negative and question markers, Steele 1977, 1995), Serrano (second position subject clitics, second position aux clitics, Steele 1977, Hill 2005), Gabriélino (second position subject clitics, Munro 2000)

**Tubatulabal** (second position subject clitics, second position auxiliary clitics, Steele 1977, Hill 2005)

**Southern Uto-Aztecan languages**

**Taracahitic languages**

Mayo (second position subject clitics, Collard and Scott 1974), Tarahumara (second position subject clitics, Steele 1977), Yaqui (second position subject clitics, Steele 1977, Dedrick and Casad 1999)

**Tepiman languages**

Pima (second position subject clitics, Munro 2000), Tepehuan (Willet 1991), Tohono O’odham/Papago (second position subject clitics, second position auxiliary clitics, Steele 1977, Hill 2005)

**Corachol languages**

Cora (second position subject clitics, Steele 1977, Haugen 2007, Langacker 1984)
Illustration: Comanche (2nd position subject clitics). Steele (1977): there is a diachronic process regarding independent pronouns and subject clitics, the latter being derived from the former. When this happens they (subject clitics) occur in the second position. (DM is a discourse marker (for topicalization)).

(21) a. tiasi-se  ni  tihka
     again-DM I eat
     ‘Again I ate.’

b. * ni  tihka
     I eat
     ‘I ate.’

c. tihka  ni
     eat I
     ‘I ate.’

(22) a. i-H/pu=u       tihiya kati-mi?a- ti=
     here-pu=he horse sit(SG SUBJ)-go-GEN:ASP
     ‘He’s riding along on a horse, going this way. Or he’s going this way, riding along on a horse.’

b. nah  utii=hi=pe-H/tu=ni-wiHtuʔi-ka= tuʔi
     just they when=H/tu=my-wait=for=someone-??-UR:ASP
     ‘They doubt if I will be ready.’

(23) a. tihka  ni
     eat I
     ‘I ate.’

b. tahi-taʔo-ʔai-ki=i  nii
     us=DU=INCL-pound=meat=make-BEN=CMPL:ASP I
     ‘I made pound of meat for the two of us.’

c. ke  nii  toHtin-kaHtu=miʔa-wai-tf
     NEG I name-toward go-wai-GEN:ASP
     ‘I will not go to Lawton.’

The subject is a clitic located in the second position, the V can precede it or follow it. It does not have to be V-adjacent (22), and either one word or a full phrase (VP in (23)b) can precede it, as (23) shows.

(i) nihi-ti i tuʔa
     us=DU=EXCL help
     ‘Help us!’

(ii) tahi-taʔo-ʔai-ki=i  nii
     us=DU=INCL-pound=meat=make-BEN=CMPL:ASP I
     ‘I made pound of meat for the two of us.’

The DP/NP status of the Uto-Aztecan languages in question:
Most of them are clearly NP languages and in fact do not have a definite article, e.g. Comanche. The literature occasionally cites some of these languages (in particular, Southern Paiute, Cupeño, Tohono O’odham, Yaqui, and Cora) as having articles.
There is no form that only functions as a definite article in Southern Paiute. The form that is sometimes considered to be a definite article, -u’, is a demonstrative (see Givón 2011). It is also not obligatory for definite interpretation (see Givón 2011, Shopen 2007).

The same holds for Cupeño pe’ (see Hill 2005) and Yaqui u (see Guerrero 2004 (e.g. p. 20), Guerrero and Belloro 2010 (e.g. p. 118 and 121), Dedrick and Casad 1999 (e.g. p. 68 and 193)).

Tohono O’odham: the form that is sometimes claimed to be a definite article, g, can be apparently used either as a definite or an indefinite article, it can be used without a noun, and is not required for definite interpretation (see Zepeda 1983). It is also not distinct from a demonstrative (see Mason 1950).

The same holds for Cora, which I will use to illustrate these issues. Forms that are sometimes translated as definite articles (Casad 1984), like ART below, are in fact demonstratives. They also do not obligatorily result in definite interpretation (compare the first and the second ART in (24)), and are not required for definite interpretation (25).

(24) an-ká-cu'u-ta'i-ri'-i
   on.top-down-break-burn-make-STAT ART spoon ART fire with
   ‘The edge of the head of the spoon is burned off by a fire.’ (Casad 1984:191)

(25)    ka-nú=r-áh-ča'i
       NEG-I=DISTR:SG-(?)-have soap
       ‘I don’t have the soap.’ (Casad 1984: 188)

Uto-Aztecan languages confirm (5).

**Conclusion:** among the following fifty-two languages with 2nd clitics there are no counterexamples to (2)/(5): Serbo-Croatian, Czech, Slovak, Slovenian, Hucul Ukrainian, Sorbian, Latin, Ancient Greek, Hittite, Sanskrit, Old English, Ossetic, Northern Talysh, Southern Tati, Pashto, Tagalog, Yingkarta, Wajarri, Ngiyamba, Warlpiri, Warumungu, Wambaya, Garrwa, Pitjantjatjara, Yir-Yoront, Yukulta, Nhandu, Gurindji, Djaru, Ngarinyinman, Mudburra, Wembawemba, Wergaia, Madimadi, Wathawurrung, Woiwurrung, Bilinarra, Warmman, Comanche, Chemehuevi, Southern Paiute/Ute, Cupeño, Luiseño, Serrano, Gabrielson, Tubululabal, Mayo, Yakui, Pima, Tepehuan, Tohono O’odham/Papago, and Cora.

**The DP/NP difference more broadly**

**Brief illustrations of the differences**

**Extraction out of the nominal domain**

What is good in English is bad in SC, what is bad in SC is good in English

-Adjectival modifier of a noun: does not extract in English, extracts in SC

-Complement of N (of-genitive in English, genitive in SC): extracts in English, does not extract in SC

-PP-adjunct modifier (non-complement) of an NP: does not extract in English, extracts in SC

Locality of extraction out of the nominal domain is completely different in English and SC.

We have two things to work with here: (a) structural differences (b) the locality system itself—phases

I argue for (a) (Bošković 2013, 2014). Assuming uniform structure leads to positing parameterization with respect to phases.

**Word order**

Word order in the nominal domain is generally freer in languages without articles. E.g. demonstratives, possessives, and adjectives can all co-occur in Chinese, any order is in principle possible. Depending on the meaning, the demonstrative occurs in different positions in SC (see Bošković 2016b on the latter).

Richer structure imposes syntactic constraints on word order (e.g. No DP to impose syntactic constraints on word order in article-less languages; no DP to force demonstratives into a unique position)
**Binding**
The possessor in examples like *John’s book* binds out of the TNP in SC, not in English. Binding of reflexive possessors and reflexives in general also works differently (see Despić 2011, 2015 on this; see also Franks 2017, who shows that La Terza’s 2016 argument against Despić is empirically flawed).

**Generalizations**
Bošković (2008a, 2012): there is a number of crosslinguistic generalizations where languages differ with respect to a number of syntactic and semantic phenomena depending on whether or not they have articles, which means that the presence or absence of articles cannot simply be a phonological (i.e. PF) effect. A selection of these generalizations is given in (26).

(26) **NP/DP generalizations** (see Bošković 2008a, 2012 and references therein)
1. Only languages without articles may allow left-branch extraction out of TNPs.
2. Only languages without articles may allow adjunct extraction from TNPs.
3. Only languages without articles may allow scrambling.
4. Multiple-wh fronting languages without articles do not show superiority effects.
5. Only languages with articles may allow clitic doubling.
6. Head-internal relatives display island sensitivity in languages without articles, but not in languages with articles.
7. Polysynthetic languages do not have articles.
8. Only languages with articles allow the majority reading of most.
9. Languages without articles disallow negative raising (i.e. strict clause-mate NPI licensing under negative raising); those with articles allow it.
10. Negative constituents must be marked for focus in article-less languages.
11. The negative concord reading may be absent with multiple complex negative constituents only in negative concord languages with articles.
12. Number morphology may not be obligatory only in TNPs of languages without articles.
13. Radical pro-drop may be possible only in languages without articles.
14. Elements undergoing focus movement are subject to a verb adjacency requirement only in languages with articles.
15. Inverse scope for S-O is unavailable in languages without articles.
16. Possessors may induce an exhaustivity presupposition only in languages with articles.
17. The sequence of Tense phenomenon is found only in languages with articles.
18. Second position clitic systems are found only in languages without articles.
19. Obligatory numeral classifier systems are found only in languages without articles.
20. Only languages without articles may allow subject reflexives.

(Some illustrations: Left branch extraction of adjectival elements)

(27) *Expensive he saw [t; cars]

(28) Doroğuşu on videl [t; mašinu] (Russian)
expensive he saw car


(29) Only languages without articles may allow LB examples like (27).

Bošković (2012): Bulgarian and Macedonian vs other Slavic languages
Latin vs Modern Romance
Mohawk, Southern Tiwa, Gunwinjguan languages (see Baker 1996), Hindi, Bangla, Angika, and Magahi also allow LB and lack articles. Coll. Finnish has developed an article; LB allowed only in literary Finnish, no article there (Franks 2007)

(30) a. Punaisen ostin auton. [literary Finnish, poetic style]
    red-acc buy-pst-1sg car-acc
b. *Punaisen ostin (sen) auton. [spoken Finnish]
    red-acc buy-pst-1sg the car-acc

History of Greek (Bošković 2012 based on Taylor 1990)
Homeric Greek (8th century BC, Iliad and Odyssey) was an article-less language, Koine Greek (1st century AD, the New Testament corpus) was a full-blown article language
Homeric Greek productively allowed LBE, Koine Greek did not.

Adjunct extraction from TNP

(31) a. Peter met [NP girls from this city] b. *From which city did Peter meet [NP girls t_{i}]

Stjepanović (1998), Bošković (2012): SC and Russian, which have no articles and allow LB, allow extraction of adjuncts out of NP (the same holds for Czech, Polish, Ukrainian, Slovenian, Hindi, Bangla, Angika, and Magahi); Bulgarian, which has articles and does not allow LB, does not (the same holds for Spanish, Icelandic, Dutch, German, Arabic, and Basque).

(32) Iz kojeg grada je Petar sreo [djevojke t_{i}] (SC)
    from which city is Peter met girls
(33) *Ot koj gradi Petko [sreštna momičeta t_{i}]? (Bg, Stjepanović 1998)
    from which city Petko met girls
(34) *Frá hvaða borg séð þú stelpur? (Icel andic)
    from which city see you girls

(A factor to control: an adjunct in one language can be an argument in another language, see Ticio 2003)

(35) Only languages without articles may allow adjunct extraction out of TNPs.

Scrambling

(36) Only languages without articles may allow scrambling.

SC, Russian, Polish, Czech, Latin, Japanese, Korean, Turkish, Hindi, Chukchi, Chichewa, Mohawk, Warlpiri... have scrambling and lack articles (what counts is long-distance scrambling from finite clauses)
Latin vs Modern Romance Lakhota vs Mohawk and Wichita

Superiority and multiple wh-fronting

(37) a. Koj kogo vižda? b.*Kogo koj vižda?
    who whom sees
    * ‘Who sees whom?’ (Bulgarian)
(38) a. Ko koga vidi? b. Koga ko vidi?
    who whom sees
    (SC)
(39) MWF languages without articles do not display superiority effects in examples like (37)-(38).
MWF languages without articles do not show Superiority effects: SC, Polish, Czech, Russian, Slovenian, Ukrainian, Mohawk
MWF languages that show Superiority effects all have articles: Romanian, Bulgarian, Macedonian, Basque, and Yiddish. Hungarian is an exception (articles and no superiority), which doesn’t violate (39).

Superlatives
Živanović (2007):(40) doesn’t have the reading where more than half the people drink beer. It only has the reading where more people drink beer than any other drink though it could be less than half the people

(40) Največ ljudi pije pivo.
most people drink beer.
‘More people drink beer than drink any other beverage.’ (Plurality reading, MR)
‘*More than half the people drink beer.’ (Majority reading, PR)

English most gives rise to both readings, though in different contexts. German MOST has both readings.

(41) Die meisten Leute trinken Bier.
the most people drink beer.
‘More than half the people drink beer’/‘More people drink beer than any other drink’ (focus on beer)

Bošković (2012): English, German, Dutch, Hungarian, Romanian, Macedonian, Bulgarian, Basque, Arabic, which have articles, allow MR. MR is disallowed in Slovenian, Czech, Polish, SC, Chinese, Turkish, Hindi, Angika, Magahi, and Punjabi, which lack articles and allow only the plurality reading. We then have (42) (I set aside cases where the majority reading is expressed with a noun like majority).

(42) Only languages with articles allow the majority superlative reading.

Polysynthetic languages

(43) Polysynthetic languages do not have articles

Classifiers Cheng (2013)

(44) Obligatory nominal classifier systems are found only in languages without articles

Many additional phenomena either work differently in languages with and without articles, or they can be present only in one of these (the former: focus movement, scope, number morphology, head internal relatives, negative constituents, interpretation of possessors; the latter clitic doubling, sequence of Tense, negative raising, radical pro-drop, subject reflexives, see the generalizations in (26))

Language acquisition
The NP/DP generalizations all involve potential triggers but most of them (even all of them) are not plausible candidates
How about the definite article?
Are there any DP languages with a null definite article? I.e. do all languages without an overt definite article lack DP?
This seems to be the case. So, definite article is in principle a perfect trigger.
In languages with articles, children do omit articles/D-elements early on; proposals that children go through the NP stage, which would then be a default (Guasti, Gavarro, de Lange and Caprin 2008; Mathewson, Bryant, and Roeper 2001).
Koulidobrova (in press) on the emergence of D-items in the child's spontaneous speech (English). D-items emerge as a set, and their emergence in child's speech is correlated with the emergence of the definite article, i.e. the definite article predicts the emergence of DP associated items. (Koulidobrova interprets the full range of her data as supporting the DP/NP analysis; see her work)

**Why is definiteness so important?**
Definiteness is a semantic notion, what is really important here, its reflex in the syntax (+/- DP) or the semantic notion itself (i.e. semantics)?

Iota operator (which is the semantic job of definite article) turns NP, which is of type <e,t>, to type e; in other words, in turns a predicate into an argument, making it possible to integrate the NP into the clause/VP (see also Progovac 2010 for an idea that this is all there was in one stage of proto-syntax, i.e. a single argument-predicate merge)

Broad typological correlations are the key to understanding the nature of language (but we have to ask the question why they hold)

**Talić (2017): more on the NP/DP typology**

Languages with affixal articles (like Bulgarian, Romanian, Icelandic, Norwegian, Swedish, Danish) can drop the article in certain contexts and behave like languages without articles when the article is dropped

Articles in these languages are different from languages like English in their PF manifestation. In the basic cases the definite article in these languages is an affix/clitic which does not occur DP-initially where articles typically occur in languages with head-initial projections in the TNP. Since the definite article is an affix on a noun, the affix can be taken to realize a feature on the noun that needs to be licensed by a syntactic head, instead of being base generated in a separate head position (Anderssen 2007 notes affixal article in Norwegian is acquired earlier than English article, which may suggest that children treat the suffixal article in Norwegian as a realization of a feature on the noun).

Talić: DP must be present in English for a formal (feature-licensing reason). In affixal article languages DP is there when motivated by interface considerations. This means that it either has to have overt PF manifestation or that it is required by semantics.

The semantic contribution of the: the definite article picks out an individual from a set (e.g. Chierchia 1998: the turns predicates to individuals)
Chierchia 1998: Languages without articles have semantic type-shifting operations that pick out individuals without requiring the presence of D in the syntax.
Such operations are not available in languages that have articles.
Thus, in e.g. Bulgarian, DP is required for interpretation since it has a definite article.

→ Two different motivations for DP:
  - Deeply DP languages, e.g. English → Deep formal considerations.
  - Not-so-Deeply DP languages, e.g. Bulgarian (when motivated by interface considerations-PF manifestation and when required by semantics)

The definite article in superlatives does not contribute the definiteness interpretation it has in non-superlative contexts.
Extraction out of definite DPs is degraded in English (definiteness effect). Superlative DPs, despite the presence of the definite article, do not induce the definiteness effect.

(45) a. Who$_i$ did you see pictures/a picture of $t_i$?
   b. *Who$_i$ did you see the/these pictures of $t_i$?

(46) Who$_i$ did you see the best picture of $t_i$?

In superlatives uniqueness is imposed by the semantics of the –est morpheme (e.g. Sharvit&Stateva 2002). *The in English superlatives only indicates that there is a DP there, it doesn’t contribute the uniqueness interpretation.

The article can be omitted with superlatives in languages with affixal articles (The interpretation of (a) and (b) is different, (b) patterning with the superlatives in NP languages in this respect, see Zheng 2016)

(47) a. Ivan ima naj-dobri-te albumi ot U2.
   Ivan has superlative-good-the albums by U2.
   b. Ivan ima naj-dobri albumi ot U2.
   Ivan has superlative-good albums by U2 (Bulgarian)

(48) John has the best/*best albums by U2.

The majority reading of ‘most’, which is allowed only in languages without articles (42).

(49) a. Die meisten Leute trinken Bier.
   the most people drink beer
   b. ‘More people drink beer than any other beverage’ (Plurality reading) (focus on beer)
   c. ‘More than half the people are drinking beer.’ (Majority reading)

Dubinsky and Tasseva-Kurtchieva (2014): in Bulgarian the majority reading of most depends on the presence/absence of the definite article: it is disallowed in (50)a, but allowed in (50)b.

(50)a. Poveče hora poznavat Ivan.
   more people know Ivan
   b. Poveče to hora poznavat Ivan.
   more.the people know Ivan

Adjunct extraction is not possible in Bulgarian when the article is present, but it is when the article is dropped (Dubinsky and Tasseva-Kukchtieva 2014).

(51) a.*[Ot koj universitet]$_t$ srešna-ha nyakolko-to studenti $t_i$?
   from which university met-they several-the students
   ‘From which university did they meet several students $t$?’
   b. [Ot koj universitet]$_t$ srešna-ha nyakolko studenti $t_i$?
   from which university met-they several students

Weak definites; Aguilar-Guevara (2014): the definite article in weak definites in English lacks its prototypical interpretation that involves familiarity presupposition.

(52) She went to the dentist.

Icelandic, Bulgarian, and Romanian can often omit the definite article in this kind of contexts.
(53) a. Hún fór til tannlæknis.
    she went to dentist
    ‘She went to the dentist.’
b. Hann fór út í búð.
    he went out in store
    ‘He went to the store.’
c. Ég tók rútu í skóla-nn.
    I took bus in school-the
    ‘I took the bus to school all my life.’
d. (Tja) otide na zbolekar.
    (she) went to dentist
    ‘She went to the dentist.’
f. Cjal jivot pătuvaḥ s avtobus
    whole life traveled with bus
    ‘I travelled with the bus all of my life.’
g. S-a dus la pravalie.
    REFL-has went to store
    ‘He went to the store.

For additional cases where languages with affixal articles pattern with languages without articles, see Talić (2017), Bošković (2008b) (regarding wh-islands), Reuland (2011), Despić (2011, 2015) (regarding anaphor binding), and Oda (2018) (regarding extraction out of conjuncts) In a number of cases, it is possible to drop the definite article in affixal article languages when there is no semantic motivation for it, which Talić interprets as indicating that affixal article languages can lack the DP layer in the TNP when its presence has neither semantic motivation nor phonological manifestation. Articles are still needed in most cases to contribute the right semantic interpretation of TNPs in these languages so the DP is usually projected. In languages like English, articles are needed for formal reasons.

**Back to second position clitics**

Clitic second is not structural in nature.

Second position clitics do not occur in the same head position (see Bošković 2001 and references therein for a number of additional arguments).

Splitting the clitic cluster (relevant clitics are given in italics)

(67) Mi smo mu ga dali, a i vi ste (?)mu) ga—dalı (takodje).
    we are him.dat it.acc given and also you are him.dat it.acc given too
    ‘We gave it to him, and you did too.’ (Stjepanović 1998)

(68) Ivan je [VP kupio auto] i [VP razbio ga]
    Ivan is bought car and ruined it
    ‘Ivan bought a car and ruined it.’ (Wilder and Ćavarić 1997)

(69) shows that even clause-mate clitics can be separated as long as the intervening material is a full intonational phrase so that each clitic is second in its intonational phrase.

(69) Oni su, kao što sam vam rekla, predstavili se Petru.
    they are am you.dat said introduced self.acc Petar.dat
    ‘They, as I told you, introduced themselves to Petar.’ (Bošković 2001)
(70a) shows the aux and the ethical dative clitic can occur above sentential adverbs, which is not possible with argumental dative (and accusative) clitics (70b), showing they don’t all occur in the same position.

(70) a. Oni su ti pravilno odgovorili Mileni. (ti=ethical dative)  
    they are you.dat correctly answered Milena.dat  
    ‘They did the right thing in answering Milena.’  
    ‘They gave Milena a correct answer.’  

b. Oni su joj pravilno odgovorili.  
    they are her.dat correctly answered  
    ‘*They did the right thing in answering her.’  
    ‘They gave her a correct answer.’  

(Bošković 2001)

Talić (2018): an accent shift involving clitics in a dialect spoken in Bosnia and Herzegovina which has a locality requirement where the clitic and the host must be in the same phrase. (71) give the tones that arise as a result of high tone spread from the enclitic (in bold) to its host. (71) indicates that the wh-C enclitic li and the aux enclitic in questions are located within CP, hence trigger high tone spread, but aux clitics in declaratives and pronominal clitics are not in CP, hence they do not trigger high tone spread.

(71) a. Štá li hoče?  
    what Q wants  
    ‘I wonder what he wants.’

b. Štá su rekli?  
    what are said  
    ‘What did they say?’

c. Šta mu govori?  
    what him.dat says  
    ‘What is (s)he telling him?’

d. da su mu govorili  
    that are him.dat said  
    ‘that they were telling him’  

(Talić 2018)

None of the operations that can split a clitic cluster in SC are possible in Bulgarian/Mac (see Bošković 2001), where the clitic cluster is inseparable (it also cannot be separated from the verb by non-clitics).

(72) a. *Nie sme mu go dali, i vie ste mu (go) dali (sūšto).  
    we are him.dat it.acc given and you are him.dat it.acc  
    ‘We gave it to him, and you did too.’

b. *Te sa, kakto ti kazax, predstavili se na Petūr.  
    they are as you.dat told introduced self.acc to Peter  
    ‘They have, as I told you, introduced themselves to Peter.’  

(Bošković 2001)

Bošković (2001) these differences indicate that SC clitics are located in separate projections--they do not all cluster in the same head position, while Bulgarian clitics do cluster in the same head position. As a result, SC clitics can be split, while Bulgarian clitics cannot be.

The correct statement of the clitic second effect in SC is actually not syntactic, but prosodic (Bošković (2001) and Radanović-Kocić (1988))

(73) SC clitics occur in the second position of their intonational (I-) phrase.

(74) #Oni su, #kao što sam vam rekla, #predstavili se Petru.  
    they are as am you.dat said introduced self.acc Petar.dat  
    ‘They, as I told you, introduced themselves to Petar.’  

(Bošković 2001)

Additional illustrations: the delayer, which brings in an additional I-phrase, is a heavy fronted constituent in (75a), a parenthetical in (75b), and an appositive relative in (75c): the clitics are located in the second position of their I-phrase (the delayers are parsed as separate I-phrases, see e.g. Nespor & Vogel 1986, Selkirk 1986, Hayes 1989).
(75) a. Sa Petrom Petrovićem #srela se samo Milena.
    with Peter Petrović met self only Milena
    ‘With Peter Petrović, only Milena met.’
    b. Znači da, kao što rekoh, #oni će sutra doći.
    means that as said they will tomorrow arrive
    ‘It means that, as I said, they will arrive tomorrow.’
    c. Ja, tvoja mama, #obećala sam ti sladoled.
    ‘I, your mother, promised you an ice cream.’ (Bošković 2001)

Bošković (in press) on Verb-second: no structural unification with clitic second (in fact, no correlation between articles and V-2). However, the same prosodic effect is found with V-2.

(76) Wie reich sie auch sei,# ich heiratete sie nicht.
    however rich she too may-be I would-marry her not
    ‘However rich she may be, I would not marry her.’  (Boeckx 1998)

It is also not possible to have a pause in front of the verb in V-2 constructions. While English however is typically followed by a pause, emellertid in (77) cannot be followed by a pause.

(77) Emellertid kan du inte använda en DVD-RAM skiva som startskiva.
    however can you not use a DVD-RAM disc as a start-up disc
    ‘However, you cannot use a DVD-RAM disc as a start-up disc.’ (Swedish, Holmberg 2015)

Swedish (78): the prosodic relationship between the adverbials themselves, and the last adverbial and the verb, is different. The adverbials are separated by a comma break from each other, indicating that an adverbial that is followed by an adverbial forms an I-phrase in (78). While due to the high pitch at the end of the adverbial there can be a sharp drop between the final adverbial and the verb, a comma break is not possible here: there is an I-phrase boundary after the adverbials that are followed by an adverbial and a phonological phrase boundary after the adverbial that is followed by the verb. This makes mötte second in its I-phrase in (78). Not possible to have an I-phrase boundary before never in (79).

(78) I går, vid femtiden, utanför stationen, när jag kom från jobbet, mötte jag en gammal skolkamrat.
    yesterday at about five outside the station, when I came from work, met an old schoolmate

(79) *Nede vid ån, under bron, tydligen aldrig har det bott en bisamråtta.
    down by the river, under the bridge, apparently never has there lived a muskrat
    (Swedish, Holmberg 2015)

In early Indo-European, finite verbs in main clauses were accentless second position elements (see Wackernagel 1892). Germanic V-2 effect may then to some extent be a remnant of the more general clitic second requirement on verbs in early Indo-European (though the effect is no longer confined to accentless verbs; see also Bošković in press on Northern Norwegian).\textsuperscript{8}

\textsuperscript{8}Wackernagel in fact suggested that finite verb cliticization led to the development of verb second (where verb second began with mono and disyllabic verbs, getting extended to longer forms); for relevant discussion see also Hock 1991, Anderson 1993, Kuhn 1933, Suzuki 2008, among others). Wackernagel also traced back modern German V-2 to Proto-Indo-European, where finite verbs cliticized to the clause-initial word in main clauses.
See also Bošković (in press) for a prosodic explanation why adjacency of the host to an I-phrase boundary matters, which unifies this with Japanese Case-marker stranding ellipsis, where the Case-marker is exceptionally stressed when adjacent to an I-phrase boundary (see Shibata 2014, Bošković 2015).  

(80) Naomi-mo moo tsuki-masi-ta ka? Naomi-GA mada tsuki-mase-n
    Naomi-also already arrive-pol-past q Naomi-nom yet arrive-pol-neg
    ‘Has Naomi already arrived? She has not arrived yet.’ (Japanese, Otaki 2011)

See also Bošković (in press) for a perspective on V-2 where the prosodic requirement (supporting a distressed verb) got grammaticalized by turning into an actual syntactic requirement where the prosodic second requirement has undergone a change to a syntactic requirement to have a Spec (which also explains the non-pickiness of V-2, where, in contrast to the usual situation where a head is picky regarding the kind of element that satisfies its Spec requirement, just about anything can satisfy V-2).

Under this perspective much of the variation among V-2 languages regarding exceptional non-V-2 cases may come from prosody (it’s well-known that there is variation both across languages and individual lexical items of a single language regarding their prosodic properties, including intonational phrasing), differences in the extent to which the V-2 requirement got grammaticalized providing another source of variation here. Any syntactic account of V-2 should include a prosody-based module.

Bošković (in press) on the development of the second position effect: the original Indo-European 2nd position effect with Vs as a combined syntax/prosody effect, with the V located in the left periphery and constructions where the syntax leaves it in a position where it is not 2nd in its I-phrase filtered out in PF.

(81) Second position requirement
      / \
      (S) Syntactic (V-in-C⁰)   (P) Prosodic (enclitic/adjacency to an I-phrase boundary)

SC clitics, which can occur quite low in the structure, only have P (i.e. they are only subject to P from (81)). Germanic V-2 involves S (hence it involves V-movement into the left periphery), with the remnant of P, which got grammaticalized in many cases.

Relevance of word order: Early Indo-European languages had way more freedom of word order than modern Germanic. What was responsible for this is a difference in the availability of Japanese-style scrambling (JSS) (which is different from what is referred to as scrambling in Germanic, see Bošković 2004). The loss of JSS has made it more difficult for the relevant element to be subject to S (i.e. to be located high in the structure) and still satisfy P given the unavailability of JSS, which could “accidentally” satisfy P in proto-Indo-European. This then led to the grammaticalization of P (in terms of a non-picky EPP requirement (without Agree, see Bošković in press)).

Dadan (in preparation): more on diachronic change (see also Dadan 2019)
Diachronic change often involves loss of movement (loss of obligatory wh-movement from Old Japanese to modern Japanese (Ikawa 1998, Watanabe 2002), from archaic to modern Chinese (Aldridge 2010),

---
9The Japanese Case-marker stranding, where the Case-marker is exceptionally stressed, is a matrix phenomenon. Early V-2, where the verb was exceptionally distressed, was also confined to main clauses (see Wackernagel 1892, Kuhn 1933). What is relevant here is that in these case the I-phrase boundary is stronger—it is also an utterance boundary (see Bošković in press).
10Prosodic variation is more likely to be involved in V-3 cases, and the extent of grammaticalization) in V-1 cases.
11One of the NP/DP generalization concerns JSS here, which is available only in languages without articles.
from Vedic Sanskrit to modern Indic languages (Hale 1987), or from Latin (Spevak 2010, Danckaert 2012, Ledgeway 2012) to modern Romance (Reglero 2004) (wh-in-situ possible in modern Romance), on-going change in Navarro-Labourdin Basque, loss of V-2 (e.g. Old Romance, Wolfe 2018, English).

This is actually loss of a specifier. Another way to lose a Spec Bošković (2001): Serbo-Croatian Q/focus marker -li has lost its ability to support a specifier; it cannot host unambiguously phrasal elements (82)a or license sluicing, which requires a Spec-head relation (82)b.

(82) a. Koga li (Petar) voli?
   whom LI Peter loves
   ‘Who is it that Peter loves?’

b. *Čiju ženu li (Petar) voli?
   whose wife LI Peter loves?
   ‘Whose wife does Peter love?’

c. Vidi nekoga. *Kogo li vidi?
   sees someone whom LI sees
   ‘He sees someone. Who?’

(Bošković 2001:27)

Bulgarian allows both full phrases and ellipsis with li.

(83) Novata kola li prodade Petko?
   new-the car Q sold Petko
   ‘Did Petko sell the expensive car?’

(84) Novata küšta li? Kogo li?
   New-the house Q whom Q
   ‘The new house? ‘Whom?’

(Bošković 2001)

Another way of losing specifiers is reanalyzing them as heads: especially prolific in the domain of complementizers, where phrases (especially specifiers of embedded CP) get reanalyzed as complementizer heads.

Georgian: interrogative wh-phrase ray ‘what’ > complementizer raytamca (Harris and Campbell, 1995); Russian and Bulgarian čto.INSTR ‘what’> čem ‘than’ (comparison complementizer); Bulgarian: ‘than how much’ (ot-kolko-to?)> otkolkoto ‘than’ (Willis 2007); English how > subordinating complementizer head (Huddleston and Pullum 2002) (also many Slavic languages, e.g. Polish, Slovak jak or Breton penaos); English relative marker that (from specifier of CP) (van Gelderen 2004); French par ce que ‘by this that’ > parce que ‘because’, etc. (van Gelderen 2004); French par ce que ‘by this that’ > parce que ‘because’, of Early Germanic hwœt reanalyzed as a C-head in exclamatives (Walkden 2014).

The emergence of agreeing complementizers from pronouns in Welsh, e.g. mi deriving from 1SG subject pronoun, and the particle fe from a masculine 3SG subject pronoun (Willis 2007) (what facilitated this was pronoun doubling, as in (86).

(85) Mi welais I ‘r gêm
   PRT see.PAST.1SG I the game

(86) Mi arhosais (.) fi
   1SG.IND wait.PAST.1SG 1SG.IND
   ‘I waited, me.’ (Willis 2007: 459)

The way structure building works favors head-complement relations, which involve merger of a head and a phrase, over traditional Spec-head relations, which involve merger of two phrases (at the point of merger). Essentially, merger of two phrases requires an additional step to label the object in question,
which is not the case with the merger of a head and a phrase, as in head complement cases (Chomsky 2013, Bošković 2016c)

Wh-movement in Latin/modern Romance: multiple wh-fronting — single wh-fronting — wh-in-situ emerging
Multiple movement harder to lose than single movement for formal reasons related to the driving force of movement (whether it resides in the target, as with single movement, or the moving elements, as in multiple movement, see Bošković 1999)

Why doesn’t all movement gets lost: essentially functional, and more broadly non-syntactic reasons
Dadan: Diachronically, the common pattern is that we observe the loss of movement instead of its gain. Any case of gaining specifiers cannot be syntax-driven, must be interface-driven and should be attributed to extra-syntactic factors, e.g. prosody or pragmatics.

Similar pattern in language acquisition: acquisition of the reflexive clitics się (SE-reflexive) in Polish
Omission of SE-clitics in contexts where they are obligatory in adult Polish (anti-causatives (87)a; body-grooming (87)b, or inherently marked reflexiva tantum (87)c:

(87) a. Zepsuła *(się) lampa  b. Co *(się) kąpie  c. pogniewała *(się)
    broke       SE      lamp     which SE     bathes       got-angry       SE
    Int: ‘The lamp broke’     Int: ‘The one who is bathing’     Int: ‘She got angry’
    (Basia 2;0; Szuman; bas200) (Kasia 1;10, Szuman corp. kas110) (Wawrzon 2;6, Weist-Jarosz waw09)

Following Kayne (1975), Marantz (1984), McGinnis (2004), where these elements are generated as the external argument, with the lower arguments raising to subject position, the omission of się in child language is in fact the instance of avoiding creation of the specifier (here, specifier of vP).

Todorović (2016): more on the correspondence between morphology and the syntactic structure

(88) TP is absent in languages that lack overt temporal morphology, i.e., TP must be realized by overt temporal morphology.

There are structural differences between languages with and without temporal morphology — the former involve a richer syntactic structure (specifically, the presence of a TP) than the latter (see also Migdalski 2015 on the presence/absence of TP). Temporal interpretations can be achieved through either Tense or Aspect, which means either traditional tense-dedicated or aspect morphology. Absence of temporal morphology in a language leads to rich aspectual morphology, which is needed to express temporal relations. Given the tendency to minimize redundancy, languages with rich aspectual morphology in fact tend not to have pure temporal morphology.

Some languages without TP: Serbian (rich verbal morphology, but traditional temporal-dedicated morphology actually denotes agreement markers), Slovene, Czech, Slovak, Russian, Polish, Chinese, Korean, Paraguayan Guaraní, Hausa, Kalaallisut, Yukatek Maya, Halkomelem Salish, Lilooet Salish, Turkish, Korean. These languages pattern with respect to a number of syntactic and semantic properties pertaining to Tense and Aspect.

One of Todorović’s tests: a particular type of mismatch between verbal forms involved in VP-ellipsis is possible only in languages that lack overt temporal morphology.

In the absence of TP, temporal interpretation can be achieved with the help of aspectual and modal components. What has traditionally been analyzed as tenses in Serbian can receive a range of interpretations which are otherwise puzzling if Tense is present in the structure of these forms.
Don’t trust traditional labels: Aorist and Imperfectum observe aspectual restrictions in Serbian (aorist occurs only with perfective aspect (89)a, and imperfectum only with imperfective aspect (89)b. They don’t in Bulgarian (which also does not tolerate finiteness mismatches under VP ellipsis and has temporal morphology, see Scatton 1984): both aorist (90) and imperfectum (90) can occur with either imperfective or perfective aspect (the difference in meaning between the two forms in (89)-(90) is crucially contributed by aspect, e.g. whether the emphasis is placed on the completion, or the lack thereof).

\[(89)\]
\[
\begin{align*}
\text{a. } & \text{Stiže } / \text{*stiza} \quad \text{Jovan!} \\
& \text{arrive-}pf.AOR / \text{*arrive-}impf.AOR \quad \text{Jovan} \\
& \text{‘Jovan arrived!’/*‘Jovan was arriving!’} \\
\text{b. } & \text{Oni pecijahu } / \text{*ispecijahu hleb.} \\
& \text{they bake-}impf.IM/ \text{ bake-}pf.IM \quad \text{bread} \\
& \text{‘They used to bake bread.’/*‘They used to finish baking bread.’}
\end{align*}
\]

\[(90)\]
\[
\begin{align*}
\text{a. } & \text{Včera pročetoh edna kniga.} \\
& \text{yesterday read-}pf.AOR.1sg \quad \text{one book} \\
& \text{‘Yesterday I read a book (and finished it).’} \\
\text{b. } & \text{Včera četoh edna kniga.} \\
& \text{yesterday read-}impf.AOR.1sg \quad \text{one book} \\
& \text{‘Yesterday I was reading a book.’} \\
\text{c. } & \text{Sedeše na čardaka.} \\
& \text{sit-}impf.IM.1sg \quad \text{on verandah} \\
& \text{‘He was sitting/ he used to sit on the verandah.’} \\
\text{d. } & \text{Večer sedneše na čardaka.} \\
& \text{evening sit-}pf.IM.1sg \quad \text{on verandah} \\
& \text{‘In the evening he would sit down on the verandah.’}
\end{align*}
\]

**Messick (2017): on attitude reports** (also Messick 2016)

Variation in how languages express *de se* attitude reports in finite clauses: English and other Indo-European languages do not distinguish *de se* and *de re* attitudes morphologically. (91) can be used to report an attitude with the attitude holder fully aware that said attitude is about himself (*de se*) or unaware that the attitude is about himself (*de re*). (91) can report either the scenario in (92)a or (92)b.

\[(91)\]
\[
\text{John said that he is smart.}
\]

\[(92)\]
\[
\begin{align*}
\text{a. } & \text{John said, “I am smart.”} \\
\text{b. } & \text{John, said, “he is smart.”}
\end{align*}
\]

In many languages, *de se* attitude reports are expressed via *indexical shift* where the first person pronoun is used to refer to the attitude holder (so we would have “John said that I am rich”)

\[(93)\]
\[
\begin{align*}
\text{Hesenj} & \text{ va ke } \varepsilon zj \text{ dewletia.} \\
& \text{Hesen.OBL said that I rich.be-Pres} \\
& \text{‘Hesen said that he was rich.’} \quad \text{(Zazaki, Anand & Nevins 2004:21)}
\end{align*}
\]

Messick observes a new way that languages use to mark *de se* found in Telugu (Dravidian) and Nuer (Nilo-Saharan): a *de se* attitude report in Telugu consists of a third person pronoun controlling first person agreement on the embedded verb, it involves agreement shift (so “John said that he is rich”)

\[(94)\]
\[
\begin{align*}
\text{a. } & \text{Rani [tanu exam pass ajj-aa-n-ani]} \text{ nam-mu-}\text{\-}\text{tund.} \\
& \text{Rani he exam pass happen-Past-1SG-Comp believe-Past-F.SG} \\
& \text{‘Rani believes that she passed the exam.’} \quad \text{(Telugu, Messick 2017:12)} \\
\text{b. } & \text{John c-ε caar [jen c-η Mary nēn].}
\end{align*}
\]
Some languages, like Ewe, use a logophor. Logophoric pronouns are typically found in embedded attitude reports; they cannot be the matrix subject of an out of the blue sentence. Logophoric pronoun \( y^e \) in Ewe can be used in attitude reports (95)a), but not as the matrix subject of an out of the blue context (95)b)

\[(95)\]
\[
a. \text{kofi be } y^e\text{-dzo.} \quad \text{Kofi say LOG-leave} \\
\text{‘Kofi said that he left.’} \\
b. *y^e\text{ dzo.} \quad \text{LOG leave} \\
\text{Intended: ‘He left.’ (Ewe, Pearson 2015)}
\]

Donno Sɔ uses a logophor with first person agreement (Curly, 1994).

\[(96)\]
\[
\text{Oumar iyem jembɔ paza bolum miñ tagi} \quad \text{Oumar LOG sack.DF drop left.1SG 1SG.OBJ informed} \\
\text{‘Oumar told me that he had left without the sack.’} \quad (\text{Donno Sɔ, Curly 1994})
\]

**Typology of de se marking** with representative languages

\[(97)\]

<table>
<thead>
<tr>
<th>Language</th>
<th>de se marking</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Third person pronoun</td>
</tr>
<tr>
<td>Amharic, Zazaki</td>
<td>Indexical shift</td>
</tr>
<tr>
<td>Ewe</td>
<td>Logophor</td>
</tr>
<tr>
<td>Donno Sɔ, Tamil</td>
<td>Logophor with first person agreement</td>
</tr>
<tr>
<td>Telugu, Nuer</td>
<td>Third person pronoun with first person agreement</td>
</tr>
</tbody>
</table>

Mesick observes that there is a gap in this typology. In this hypothetical language, a \textit{de se} attitude would be expressed with a first person pronoun and third person agreement, as in (98) (this surface pattern does exist; in e.g., Golin (Papuan). However when a first person pronoun controls third person agreement in such languages, it is always interpreted as a \textit{de re} attitude about the current speaker.

\[(98)\]
\[
\text{John said I is a hero.} \\
\text{Intended: ‘John said that he_dese is a hero’}
\]

Messick (2016, 2017) proposes a comprehensive syntax/morphology/semantics analysis of the Telugu/Nuer pattern where the basic idea is that when a pronoun is interpreted \textit{de se}, it is semantically first person. The LF for \textit{de se} attitude reports for languages that have indexical shift and languages with agreement shift is the same. In languages with indexical shift, the morphology allows for those features to be spelled out as 1\textsuperscript{st} person, but in languages without indexical shift, the morphology forces the features to be spelled out as a 3\textsuperscript{rd} person pronoun. With Telugu/Nuer agreement shift, the semantic features of the pronoun are able to control agreement on the embedded verb, yielding an apparent mismatch between agreement controller and the target where it appears that the semantic interpretation of the controller is influencing the agreement target.

The crucial ingredient of the account comes from Corbett (1979, 1983), which has shown that semantic features of a nominal can control agreement (e.g. a semantically plural noun \textit{committee} can control plural agreement in British English (99)).
The formal implementation of this idea in Messick accounts not only for the Telugu/Nuer agreement shift, but for the full topology of de se marking from (97) as well the typological gap noticed by Messick.

**Stegovec (in preparation): On person restrictions** (also Stegovec 2015, 2019)
In many languages, co-occurrence of weak pronominal objects is regulated by their person specification

**The Person-Case Constraint/PCC (STRONG version).** If indirect object [IO] & direct object [DO] pronominal markers co-occur, DO cannot be 1P or 2P.

   a. dahl-u -hu {li -lek} ✓ 1P/2P.IO ≫ 3P.DO  
      ‘They introduced him to me.’
   b. *dahl-u {-ni /-k} -lu ✗ 3P.IO ≫ 1P/2P.DO  
      ‘They introduced me/to him.’
   c. *dahl-u {-ni /-k} {-lek /-li} ✗ 2P/1P.IO ≫ 1P/2P.DO  
      ‘They introduced me/to you/me.’

**WEAK PCC.** When two pronominal markers co-occur, if IO marker is 3P then DO must be 3P.

2. Catalan
   a. *A en Josep, {me | te } li va recomenar la Mireia. ✗ 3P.IO ≫ 1/2P.DO  
      ‘She recommended me/to him.’ (Bonet 1991:178–179)
   b. a. Te m’ ha venut el mercader més important. ✓ 1P.IO ≫ 2P.DO  
      ‘The most important merchant has sold you to me.’
   c. Vi ci manderá. ✓ 2PIO ≫ 1P.DO  
      ‘S/he will send us to you (pl).’ (Anagnostopoulou 2005:203)

The PCC is typically discussed regarding internal arguments (EA). It is less known that PCC-like restrictions exist for external (EA)/internal argument (IA) combinations:

   a. griš -a {li / -lax /-le } ✓ 1/2/3.SU+3.O  
      ‘I/you/he pulled her.’
   b. * griš {-at / -an } {li / -lax /-le } ✗ 1/2/3.SU+1/2.O  
      ‘I/you/he pulled you/me.’

Stegovec unifies the two:

**EA-IA “PCC” (STRONG).** If subject [SU] & object [O] markers co-occur, O cannot be 1P/2P (3).

**EA-IA “PCC” (WEAK).** If SU & O markers co-occur, if the SU marker is 3P then the O marker must also be 3P (4).

4. Southern Tiwa, Kiowa-Tanoan (adapted from Allen and Frantz 1987:11-12):
Both strong and weak restrictions can be unified for EA-IA and IA-IA pairs, e.g. **WEAK restriction:** When two pronominal markers co-occur, if the SU/IO marker is 3P then the O/DO marker must be 3P. Apart from the WEAK EA-IA restriction, Southern Tiwa also has STRONG PCC for IAs: the DO can only be 3P whenever IO is also cross-referenced by the fused marker (5).

(5) Southern Tiwa (adapted from Allen and Frantz 1978:13-16)

a. **Tow-** wia-ban. ✓ 1P.SU ≫ 3P.IO ≫ 3P.DO
   ‘I gave **them** to **him/her**.’

b. **Bow-** wia-ban. ✓ 2P.SU ≫ 1P.IO ≫ 3P.DO
   ‘**You** gave **them** to **me**.’

c. *xxx-** wia-ban. ✗ 1P.SU ≫ 3P.IO ≫ 2P.DO
   ‘I gave **you** to **him/her**.’ (Rosen 1990: 677)

Stegovec (in press) observes PCC restrictions are actually insensitive to case type:
Evidence for case-insensitivity: REVERSE PCC patterns (henceforth P(erson) R(estrictions);
Slovenian has a baseline STANDARD PCC with clitics in a DAT-ACC order:

(6) Slovenian, Slavic (Stegovec 2015:108–9) — STRONG/WEAK PCC:

a. Mama **ti** ga bo predstavila. ✓ 2P.IO ≫ 3P.DO
   ‘Mom will introduce **him** to **you**.’

b. *Mama **mu** te bo predstavila. ✗ 3P.IO ≫ 2P.DO
   ‘Mom will introduce **you** to **him**.’

If the order of object clitics is reversed, a REVERSE PCC pattern obtains — the DO can now always be 1P/2P, whereas the person of IOs is restricted (just like (6)):

(7) a. Mama **te** mu bo predstavila. ✓ 2P.DO ≫ 3P.IO
   ‘Mom will introduce **you** to **him**.’

21
‘Mom will introduce you to him.’

b. *Mama ga ti bo predstavila. × 3P.IO ⇒ 2P.DO
   ‘Mom will introduce him to you.’ (Slovenian; Stegovec 2015:108–9)

Other examples: Zürich German (Werner 1999) and Czech (Sturgeon et al. 2012). Such patterns show:
PRs are not limited to inherent/lexical structural case configurations;
Position in the syntax, not grammatical function, matters (i.e. movement can feed PR).

**Direct-Inverse Systems**: Stegovec observes that given this, “person hierarchy effects” in Algonquian (see Hockett 1939, 1948, 1966; Goddard 1979) can be viewed as a case of the STANDARD/REVERSE PR alternation:

In Algonquian, if SU is 1P/2P and O is 3P, they must be reordered w.r.t. the base order (grammatical functions only recoverable from theme sign — also direct-inverse marker):

(8) Ojibwe/Nishnaabemwin, Algonquian (Valentine 2001:287) — WEAK EA-IA PR:

a. n- wa:bm -a: -g ✓ 1P.SU ⇒ 3P.DO⇒DIR
   1- see -1>3 -3PL
   ‘I see them.’ / *They see me.’

b. n- wa:bm -igo: -g × 3P.SU ⇒ 1P.DO⇒INV
   1-see -3>1 -3PL
   ‘They see me.’ / *I see them.’

Stegovec’s survey (which builds on Haspelmath 2004; Albizu 1997): It spans 101 languages from 23 families and 4 isolates:

**Indo-European**: 1 Spanish, 2 French, 3 Catalan, 4 Italian, 5 Romanian, 6 German, 7 Zürich German, 8 Swiss German, 9 Dutch, 10 Swedish, 11 English, 12 Icelandic, 13 Faroese, 14 Slovenian, 15 Serbo-Croatian, 16 Czech, 17 Polish, 18 Bulgarian, 19 Macedonian, 20 Greek, 21 Albanian, 22 Turkish, 23 Pashto, 24 Iron Ossetic, 25 Digor Ossetic, 26 Kashmiri; 27 Basque; 28 Uralic; 29 Finnish, 29 Hungarian, 30 Eastern Mansi, 31 Khanty (Ostyak), 32 Tundra Nenets; 33 Modern Standard Arabic, 34 Classical Arabic, 35 Cairene Arabic, 36 Maltese, 37 Senaya, 38 Christian Barwar, 39 Telkepe, 40 Migama, Barain; 41 Maasai/Maa; 42 Sambaa, 43 Haya, 44 Swahili, 45 Nyaturu/Rimi, 46 Limbum; 47 Georgian; 48 Abhkaz; 49 Hakha Chin, 50 Chepang, 51 Jyarong, 52 Nocte, 53 Tangut; 54 Kambera, 55 Manam, 56 Tagalog; 57 Yimas, 58 Manambu; 59 Monumbo; 60 Djaru, 61 Warlpiri; 62 Chukchi, 63 Koryak, 64 Alutor, 65 Itelmen; 66 Sahaptin, 67 Takelma; 68 Algonquin, 69 Blackfoot, 70 Cree, 71 Delaware, 72 Fox, 73 Mi’kmaq, 74 Ojibw/Nishnaabemwin, 75 Maniwaki Algonquin, 76 Passamaquoddy, 77 Potawatomi; 78 Southern Tiwa, 79 Picuris, 80 Arizona Tewa, 81 Kiowa; 82 Cherokee; 83 Tetelecingo Nahua; 84 Classical Nahua; 85 O’daham; 86 Zuni; 87 Tzotzil, 88 Kaqchikel; 89 Quechua; 90 Bella Coola, 91 Clallam, 92 Lummi, 93 Halkomelem, 94 Squamish, 95 Lushootseed; 96 Kutenai; 97 Koyukon, 98 Navajo; 99 Inuktut (Labrador), 100 Inuktut (South Baffin); 101 Mapudungun

PR variation (this includes both EA-IA and IA-IA PRs):
This is more patterns than previously reported in relation to PCC/PRs;
- **STRONG**: Maltese, Semitic (Borg and Azzopardi-Alexander 1997); Tundra Nenets, Samoyedic (Dalrymple and Nikolaeva 2011); Warlpiri, Ngarrkic (Hale 1973); etc. ...
- **WEAK**: Southern Tiwa, Kiowa-Tanoan (Allen et al. 1990; Rosen 1990); Swahili, Bantu (Riedel 2009); Hakha Chin, Tibeto-Burman (Peterson 1998); etc. ...
- **MIXED-1**: Czech, Slavic (Sturgeon et al. 2012); Pashto, Iranian (Roberts 2000); Hungarian, Finno-Ugric (É. Kiss 2013; Bárány 2015); Sahaptin, Penutian (Rude 1994); etc. ...
- **MIXED-2**: Potawatomi (Hockett 1939, 1948, 1966); Ojibwe (Rhodes 1990; Oxford 2014); Delaware, Algonquian (Goddard 1979); Spanish, Romance (Albizu 1997); etc. ...
- **ME-FIRST**: Romanian, Romance (Ciucivara 2006); Bosnian/Serbian/Croatian, Slavic (Runić 2013); Alutor, Chukotkan (Mel’čuk 1988).
- **YOU-FIRST**: Quechua (Myler 2016).
- **ONLY-YOU**: Bella Coola (Forrest 1994); Halkomelem (Jelinek and Demers 1983; Gerdzs 1988); Squamish, Salish (Jelinek and Demers 1983; Jacobs 1994).

PRs always have the following properties:
- the restriction always applies to the structurally lower marker (an intervention effect for Stegovec);
- The restriction either forces the lower marker to be 3P (*1P/2P), or bans the lower marker from being either specifically 1P or specifically 2P (speaker & addressee matter, see Stegovec’s work).

EA-IA and IA-IA PRs can co-exist in a language (e.g. Southern Tiwa), but crucially, not all logically possible combinations (in terms of differing PR “strength”) are attested:

<table>
<thead>
<tr>
<th>Type:</th>
<th>A (!)</th>
<th>B (!)</th>
<th>C (!)</th>
<th>D (√)</th>
<th>E (✓)</th>
<th>F (!)</th>
<th>G (√)</th>
<th>H (√)</th>
<th>I (!)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EA-IA</td>
<td>⊖</td>
<td>⊖</td>
<td>⊖</td>
<td>WEAK</td>
<td>⊖</td>
<td>WEAK</td>
<td>⊖</td>
<td>STRONG</td>
<td>STRONG</td>
</tr>
<tr>
<td>IA-IA</td>
<td>STRONG</td>
<td>WEAK</td>
<td>⊖</td>
<td>⊖</td>
<td>⊖</td>
<td>⊖</td>
<td>STRONG</td>
<td>STRONG</td>
<td>⊖</td>
</tr>
</tbody>
</table>

Table 3: Possible combinations of EA-IA-PRs and IA-IA-PRs
Patterns marked with ! involve lack of PR in the EA-IA domain, the IA-IA domain, or both.

(9) **Strength Implication Generalization.** If a language has both an EA-IA and an IA-IA PR, the IA-IA PR is never “weaker” than the EA-IA PR (or *EA-IA → IA-IA*).
- **TYPE D (WEAK + STRONG)**: Southern Tiwa, Kiowa-Tanoan (Allen et al. 1990; Rosen 1990); Blackfoot, Algonquian (Bliss 2013; Oxford 2014); Maasai, Eastern Nilotic (Payne et al. 1994; Lamoureux 2004); Chukchi, Chukotkan (Comrie 1979; Mel’čuk 1988); etc. ...
  ⇒ By far the most common pattern;
- **TYPE E (WEAK + WEAK)**: Alutor, Chukotkan (Mel’čuk 1988).
  ⇒ Only observable with “marker-demotion” repair;
- **TYPE G (STRONG + STRONG)**: Telkepe, Semitic (Kalin 2014); and Tangut, Qiangic (Kepping...
Stegovec provides a formal account which allows all attested but not unattested patterns, and which also explains why some patterns are very rare (e.g. ONLY-YOU). The account deduces (9) and also explains another generalization established by Stegovec, that a language can only have a reverse PR for internal arguments if it also has the standard one.

The gist of the account: no crosslinguistic variation in argument structure, in locality domains, and in operation Agree (i.e. the way agreement works), plus independently motivated parametric variation in the internal structure of pronouns, certain movement possibilities and the presence and properties of certain functional projections.

What is crucial in the account is the structural placement of a particular functional head, v; in particular, what is crucial is that EA is higher than v, and IEs are lower than v (so not simply an argument hierarchy).

We are all looking for generalizations, the question is then what they would follow from. A priori no formalist would exclude the possibility of a functional explanation.

All this brings us to what it means to be a “generative syntactician” investigating the nature of language these days: the empirical domain of inquiry is expanded to, in fact emphasizes, investigations of understudied languages and especially broad typological investigations.

Looking at linguistic phenomena in their totality, not compartmentalized by specific subfields (semantics, morphology, phonology as well as language change and language acquisition need to be constantly paid attention to).

References


Dadan, M. 2019. Movement that is Kept and Movement that is Lost: Movement, Adjunction, and Labeling in Diachronic Perspective. Talk at the 93rd Annual Meeting of the Linguistic Society of America.


Koulidobrova, E. in press. DP or not DP: Acquisitional evidence. Language Acquisition.
Kiss, K. 2013. The inverse agreement constraint in Uralic languages. Finno-Ugric Languages and Linguistics 2.