Simulating the development of bound person marking

Sander Lestrade

Abstract—**Two different grammaticalization pathways have been proposed in the literature for the development of verbal bound person marking. It is shown that both can be simulated successfully. This suggests that bound person marking may emerge spontaneously in language and need not be genetically encoded.**

I. Introduction

MOST languages index person features of the verbal arguments on the verb [1]. In some of the dominant languages in linguistics, i.e. German and English, this involves a (non-displaced) noun phrase or pronoun with the same role and referent in the same (narrow) clause [2, 7]. As Haspelmath argues, however, it is cross-linguistically much more common for such *co-nominals* to be optional. He distinguishes three types of bound person forms: *gramm-indexes*, probably the most familiar type, which obligatorily combine with a co-nominal; *cross-indexes*, the cross-linguistically most common type, which optionally combine with a co-nominal, and *pro-indexes*, which exclude this.[1] Examples, in the above order, are given in (1)–(3).

(1)   German [2, 9]
    a.   Ich komme-e
        I   come-1SG
        'I come.'
    b.   Elli komm-t
        Elli come-3SG
        'Elli comes.'

(2)   Georgian [2, 8]
    a.   v-Ø-č'am
        1.SUBJ-3.OBJ-eat
        'I eat it.'
    b.   me v-Ø-č'am   ma-s
        I   1.SUBJ-3.OBJ-eat it-DAT
        'I eat it.'

(3)   Standard Arabic [2, 10]
    a.   ra'ay-tu   l-kalb-a
        see-PRF-1SG-SUBJ DEF-dog-ACC
        'I saw a dog'
    b.   ra'ay-tu-bu
        see-PRF-1SG-SUBJ-3SG.M.OBJ
        'I saw it'

As it is generally accepted that gramm-indexes derive from free pronouns, the different types of index are naturally

---

[1]The traditional term *agreement* cannot distinguish properly between the first two types, hence the new terminology.

positioned on a grammaticalization cline [3] in which gramm-indexes grammaticalized furthest and pro-indexes only gave up their morphologically free status. There are two specific proposals for this development, to be discussed in more detail below: Givón argues gramm-indexes originate as anaphoric copies of topicalized noun phrases [4]; Ariel proposes the exact opposite, attenuated expressions of highly accessible event participant needing a more pronounced noun phrase copy [5]. The two accounts should be seen as complementing rather than as excluding each other, and there is typological and, to a lesser extent, corpus and historical evidence for both.

This paper will provide further evidence by modeling both proposals in a computer simulation of language evolution (Section III). For this, an initial proto-language is used in which there is no grammar whatsoever (cf. [6], [7]), allowing for the embedding of the results in the larger discussion about the evolution of language (Section IV). In the next section, the proposals of Givón and Ariel will be discussed in more detail first.

II. The development of argument indexing

Givón argues that gramm-indexes are the result of a re-analysis of an (overused) topic-shift construction, in which an older topic is re-established as such after an intervening gap by extraposing it ("The man... A woman... The man, he went..."). Once the discourse organizational function is neutralized as a result of frequent usage, the extraposition is no longer perceived and the anaphoric pronoun is reanalyzed as a gramm-index ([NP, pro V] → [NP pro-V]). Since subjects are more often topics than objects, this correctly accounts for the predominance of subject agreement in languages of the world. Also, since local persons are higher on the topic hierarchy than non-local ones, it predicts that gramm-indexes develop for speech participants mostly. According to [5, 211], however, "NP-detachments [=extrapositions] are prevalent for third person NPs", and hence the proposal mostly accounts for the development of third person gramm-indexes.

Ariel proposes a different route from free pronouns to gramm-indexes. According to her, it is the eventual (subject) argument that is the copy, which is necessary once a referential (i.e., non-anaphoric) pronoun is reduced to the extent that it fuses with the verb and is no longer perceived as a referring expression. This reduction is the result of high accessibility, accessible meanings being marked as such with more attenuated forms. As local persons are consistently highly accessible whereas third persons are only so when they are the continuing discourse topic [5, 221], the process mostly

involves expressions for the speaker and addressee.[2]

Both Givón and Ariel assume the same grammaticalization principles in which lexical items are gradually reduced both phonologically and semantically as a result of frequency and/or predictability, as a result of which they become gradually fused with neighboring words (cf. [8], [9]). They differ in the construction they assume as the grammaticalization source. For Givón, this is an anaphoric pronoun copy of an extrapositioned NP, whereas for Ariel, it is an attenuated referring pronoun with an NP (full or pronominal) copy. In the next section, both pathways will be implemented in a computer simulation of language development.

III. MODELING ARGUMENT INDEXING

A. Methodology

The development of argument indexing is simulated in WDWTW, an open-source R-package developed by the author in which the emergence of grammar can be simulated.[3] A population of agents starts out with a proto-language, i.e. a shared lexicon of referring expressions only, and a "language-ready brain" that recognizes communicative intentions and is capable of joint attention (cf. [6], [7], [11], [12]). Agents communicate about automatically generated events in their virtual world, for which they have to find an adequate wording. That is, their utterances have to be sufficiently distinctive given the situational context and the distribution of predicate roles over the event participants has to be made clear. In the absence of grammatical argument-marking strategies, agents initially have to use lexical ad-hoc solutions to mark information structure and role distribution if necessary. Using a natural language example, as the role distribution in case of a reading event in which a book and a man are involved follows from the semantics of the ingredients, nothing extra needs to be done. If the role distribution is not clear, however, the agent has to add a word specifying the predicate role of at least one of the participants if it wants to be understood (e.g. "woman man performer see", thereby assuming a grouping principle according to which the hearer will interpret together what stands together; [13]). Similarly, if the topic of conversation has been preposed, another expression referring to it can be used to mark its original position: "book, man thing read". Except for the grouping principle and the option to move the topic to first position, word order is uninformative.[4]

As shown in both examples already, the extra word specifying the role or referring to the topic need not be as specific as its host/target. Whereas the latter has to be sufficiently distinctive to pick out its referent in the situational context (and there may be as many as 30 distractor events going on), the former only has to distinguish between the arguments in the clause or between the predicate roles (of which there are only two maximally). Thus, in the role-disambiguation case, if the predicate roles differ in agentivity, which they mostly do in natural language, it suffices to mark one of the arguments as having the more agentive role. The exact interpretation then follows from integrating it with the verb's semantics: the performer of SEE is a seeer (cf. [14], [15]). Similarly, if there are two referential expressions between which an anaphoric expression has to distinguish, a general expression specifying a subset of their features will do mostly, e.g. specifying animacy only. Note that this generality of meaning allows for the usage in different situations and hence increases frequency of use.

The agents in the model do not know about books and women. Instead, their mental lexicon is modeled as a list of (randomly generated) forms with values on a number of abstract meaning dimensions (cf. the concept notion of [16]). Just like CAT for us is a concrete, living entity, an animal that is domesticated, from the family of cats, etc.; for an agent *esodad*, used in (4) below, is a 0 on the first three meaning dimensions, a 1 on the fourth, a 0.5 on the fifth, etc.[5] The vector representations do not have to form interpretable meanings for us, it is only to provide the agents with a set of meanings to talk about. Crucially, whereas the agents in the simulation only know about abstract numbers, the idea of words having to be sufficiently specific to get a meaning across is exactly the same.

Initially, all lexical items (including those for local persons) are fully specified semantically and have equally long forms of expression. Words that are frequent or predictable (because of the context or prior usage) are pronounced sloppily, which is instantiated as going back in the alphabet for the last letter of a form and deleting it altogether if this is no longer possible. Sloppy pronunciation does not lead to a change of lexical representation for the agent using the form. But if the (younger) hearer is still unsure about the form of a word because it has not used it sufficiently frequently yet, it will adapt its representation on the basis of what it hears (cf. [17]). Thus, words may *erode*. If the production effort of a word falls below the threshold of 15 (in which the first vowel and consonant of the alphabet cost 1, the second cost 2, etc.), markers are suffixed to their host. Also, words may *desemanticize*: First, they can extend their meaning range incidentally (if the context does not require a more specific description or in the absence of a better expression). Eventually, such an extension may become a standard part of a word's meaning, as a result of which it becomes more general. In the model, desemanticization involves the progressive removal of the meaning dimensions along which most variation is attested in it usage history (cf. e.g. [18] for the possibility of such within-generation change). Deletion takes place only after certain frequency thresholds have been reached. For a first dimension to be removed, a word has to be used in 1% of the utterances. This proportion grows exponentially to 30% for the last removable dimension (items being specified for at least two dimensions).

---

[2]Note that what is relevant the development according to Ariel is thus not frequency differences between persons, but accessibility ratio within persons [5, 247].

[3]A user-friendly version is still being developed and will be distributed via the CRAN archive [10]. In the mean-time, the codes are available from the author upon request.

[4]In the present set-up, that is. It is possible to switch on the `agentFirst` parameter, after which agents will interpret the first argument as the more agentive one.

[5]Both the dimensionality and distinctionality can be specified by the user, as virtually all other settings of the model.

The population starts with two agents with a shared lexicon of 999 randomly generated nouns and 499 such verbs. Agents die after 4000 utterances and procreate at the age of 3000, at which point their mixed lexicon is inherited by their two children, resetting all frequency numbers to zero and randomly modifying the meanings of those words that have not been used by the parents until then (the idea being that if words are not used frequently enough, they cannot be learned properly). For present purposes, these simplifications about the development and maintenance of a conventional lexicon seem warranted (for the feasibility of modeling this part of language evolution, cf. [19]–[21]).[6]

Two agents are randomly selected to take part in a conversation. One of the agents, the speaker, has to find a wording for one of the events, the *target event*, that suffices for the other agent, the hearer, to single it out. Words are activated in the mental lexicon of the agent on the basis of both referential match and frequency of usage. The most frequent and best matching words are considered first. As said above, the selected words need not describe their referents perfectly, the agent only has to deem them sufficiently distinctive given the situation and sentence context. If communication is successful (i.e., if the hearer selects the target event), the usage frequencies of the words produced and heard (which need not be the same) are raised by one, which increases the likelihood of being selected again. If communication fails, the numbers stay the same. After a communicative turn, a new situation is created. If the conversation continues, the speaker and hearer roles change, otherwise, two new speech participants are selected from the population. In each conversation, agents built up a common ground of things they have talked about, to which they are more likely to refer again for discourse coherence (the odds against introducing a new item, `oddsNew`, are 6:1) and which are hence more susceptible to erosion. Discourse participants are themselves standard part of the common ground (and thus inherently accessible, a la Ariel). Newborn agents up to the age of 200, if selected for a conversation, listen only, thus initially developing their own frequency distribution on the basis of the language sample they are exposed to.

B. Results

In this section, the results of a number of scenarios will be discussed. There are three model parameters that will be manipulated: `local`, `topicFirst`, and `refCheck`. By the first, agents may get the possibility to refer to themselves. (And if they can, the odds for topichood are 3:2:1 for first, second, and third person respectively.) If `topicFirst` is put on, to simulate the proposal by Givón, the topic of the utterance is moved to first position and a new word is selected to take its place. Recall that this new word can be rather

general in meaning, as it only has to be sufficiently distinctive to link to the topicalized noun rather than to the competing argument in the utterance, and does not have to distinguish between the distractor objects in the event situation. Also, note that all topics, and not just the contrastive ones, are moved and copied. The third parameter determines whether or not agents check if argument expressions are sufficiently expressive, simulating the proposal of Ariel. If their production effort falls below 10, a more pronounced copy is put in their stead. For each scenario, two lineages are run. As the results are completely equivalent, only examples of the first will be discussed. In the first lineage, all three parameters are set to false. Next, agents can talk about themselves, without moving topics and checking referential capacity, however. Then, testing Givón's proposal, topics may be moved and copied, and in a fourth lineage pair, testing Ariel, arguments are checked for referential capacity. Finally, all parameters are set to true, to see if the two proposals may indeed work together, as Ariel suggests.

First, let us see what happens if topics are not put in first position, agents cannot refer to themselves or each other, and the referential weight of expression is not considered (i.e., the base line, for present purposes; `topicFirst=F`, `local=F`, `refCheck=F`). With these settings, words may still grammaticalize, viz. into role markers. A typical example of an utterance is shown in (4), in which *V* is used to mark the verb and *U* stands for *undergoer*, i.e. the less agentive predicate role (cf. [22]). The marker *on ≻ onaline* is maximally short and specified for three meaning dimensions only. It is highly productive, as it is used in approximately 1/4 of the utterances to mark the undergoer. Bound person marking does not emerge with these settings (which is expected).

(4)   esodad metum-on teletal
        esodad metum-U teletal.V
        'Esodad teletals metum.'

If agents do have the possibility to refer to themselves, the speaker and hearer are referred to roughly 2000 times each, which is as often as all other objects (and which probably is overestimated). As before, topics are not moved and referential weight is not checked (`local=T; topicFirst=F; refCheck=F`). As local persons are frequently referred to in all sorts of situations, their expressions are likely to grammaticalize, and the words for speech participants quickly develop into "pronouns". Role markers again develop. Still, no systematic bound person marking develops as neither `topicFirst` nor `refCheck` are turned on. Thus, agents typically say things like (5):

(5)   ore satenes   asenos-od
        1   satenes.V asenos-U
        'I satenes asenos.'

In the third lineage, topics are put in first position and cross-referred to with a more general copy (adjacent to the verb; `local=T; topicFirst=T; refCheck=F`). Now, bound person marking should emerge according to [4], which indeed is the case. Roughly half of the times, the local

---

[6]As one of the reviewers correctly points out, having different generations is not strictly necessary to model the development of indexing. This is the default setting of the model, however, as it allows for comparison between generations, which is often convenient. Also it might be necessary to warrant the ecological validity of future applications of the model. But one could also give agents an eternal life and prohibit their procreation using the parameters `deathAge` and `procreationAge`.

pronoun is simply used twice (deictically as a subject, and anaphorically as an index; e.g. "you, I saw you"). But also, we find two specialized indexes, namely *-mo* for first and *-ur* for second person topics. Indexing for third person does not develop. An example is given in (6).

(6) es ona-mo ru-an
1 ona.V-1 you-U
'I ona you.'

At this point, it may be interesting to see how earlier speakers of this lineage utter the exact same meaning, in the exact same situation. This is shown in (7) (in which the generation of the agents is given between parentheses and word order has been adjusted for ease of comparison). In the model, the reanalysis from [NP, V pro] into [NP V-pro] goes hand in hand with the gradually increasing grammaticalization of the individual words:

(7) a. esidan onumam mutodan runalum anolote (5)
b. esi onud-moda runali anolom (10)
c. es ona-mo ru-anol (20)

In the fourth lineage, highly attenuated need a more elaborate copy if they can no longer refer properly (cf. [5]). To model this, if the production effort of a referential item is less then 10 it will be attached to the verb and a more specific expression will be placed in its stead (`local=T; refCheck=T; topicFirst=F`). Again, only local indexes develop. A typical result is shown in (8) (the homonymy with English *it* is accidental).

(8) it to musitis-ra
it 2 musitit.V-2
'You musitit it.'

In most well-known languages, like English and German, agreement is either with the subject or object. But in the cross-linguistic survey of verbal person marking of [23], most languages in fact index both the subject and object on the verb. An example is Hungarian, in which the cross-indexes have fused into a portmanteau morpheme: *Szeret-lek* means 'I like you', *szeret-em* means 'I like it'. Interestingly, such rich agreement systems may emerge under Ariel's proposal too. If both the subject and object expression are highly attenuated, both need extra expression. An example is shown in (9):

(9) to ti dasalunm-ra-at (39)
2 1 dasalunm.V-2-1
'I dasalunm you.'

Differently from the simulation of Givón, this time the pronominal markers and the verb indexes are systematically distinct: the copy has to be different, as the original does not suffice as a referential expression. Also differently from the previous lineage, this time something changes abruptly. Consider the utterances for the exact same meaning by previous

agents:[7]

(10) a. italosi musitis ratol (5)
b. itune musitis=rar (15)

In comparison with the utterances in (10), an extra word, viz. *to*, is inserted in (9) (for the reasons explained in Section II). Thus, whereas there first was no agreement/gramm-indexing, eventually it appears. In Givón's proposal, on the other hand, agreement could be said to be present throughout, only its instantiation became more and more grammatical.

A final difference with the implementation of the proposal by Givón is that the utterances in (9) and (10) can be mapped onto the different indexing types discussed in the introduction. The utterance in (9) corresponds to the gramm-index type, that in (10-b) to the pro-index type, and that in (10-a) to the grammaticalization source: the original free pronoun construction. (The cross-index type that is in-between the other two indexing types does not emerge, as the model doesn't allow for such optionality.)

As said in the introduction, the proposed diachronic mechanics do not exclude each other, and we do not have to choose between them. If both topics and attenuated pronouns are copied (`local=T; topicFirst=T; refCheck=T`), again indexing emerges:

(11) idoral tanuren nerusali-idi
1 tanuren nerusali.V-1
'I nerusal tanuren.'

IV. DISCUSSION

Although at first sight, it seems Ariel's proposal is preferable as its developmental history better matches the indexing typology described in the introduction (i.e., the pro, cross, and gramm type), both her and Givón's ideas may be right: Both proposals were modeled successfully in WDWTW, there are language histories that better fit Givon's proposal, and at least in the model, they can apply simultaneously indeed.

As said in the introduction, using an initial proto-language allows for the embedding of the results in the more general language-evolution discussion. Grammaticalization could be seen as a cyclic process, in which old forms and structures are replaced by new ones. From this perspective one could argue that there's an underlying systemic need for agreement markers and only their surface manifestations change. By starting from scratch, however, it is shown here that there need not be an in-built "agreement routine". Gramm-indexes (and pronouns before them) develop from full nouns, and there was no such thing as agreement initially. It seems plausible that eventually agreement/gramm-indexing is learned as such directly, and no longer results from the assessment of the deictic capacities of arguments. Crucially, such a reanalysis of

---

[7]There is an annoying bug in the model that I discovered too late to repair for this paper: Words are only suffixed to their host if they are considered markers, and in the model, referential expressions are not considered as such. Thus, although the to-be suffix *rar* in (10-b) reaches the suffix threshold, it is not suffixed to its host yet; it will be once it reaches the referential threshold too, as shown in (8). As this concerns a technical rather than theoretical error, I used a "=" in (10-b) to mark that *rar* should have been a suffix at this point already (but one that still is capable of referring properly).

the relation between argument and verb marker can be made on the basis of the above data, and therefore we do not need to assume an inherent agreement feature in language. In other words, it seems agreement can be removed from the language genome.

The model allows for manipulating all sorts of parameters, the discussion of which was precluded by reasons of space. In future research of indexing, there are various parameters that should be explored. To mention but two, the person topicality needs to be investigated as the odds of 3:2:1 may not be valid, and (the relative distance between) the suffix and referential threshold probably need more investigation.

Finally, WDWTW being a computer model under construction, there are many things that could be developed further. For example, indexing in the model does not really involve syntactic role (i.e. subject or object) indexing. At present, it is either a local person or the topic that is indexed. It will be interesting to see if eventually a grammatical abstraction over semantic role, topicality, and person can be made that comes closer to a syntactic role like subject.

V. CONCLUSION

Two different grammaticalization pathways have been proposed in the literature for the development of verbal bound person marking. It was shown how both could be modeled successfully. Thus, it was shown how indexing may emerge in a protolanguage and need not be genetically encoded.

ACKNOWLEDGMENT

REFERENCES

[1] A. Siewierska and D. Bakker, "Case and alternative strategies: word order and agreement marking," in *The Oxford handbook of case*, A. Malchukov and A. Spencer, Eds. Oxford: Oxford University Press, 2009, pp. 290–303.

[2] M. Haspelmath, "Argument indexing: A conceptual framework for the syntactic status of bound person forms," in *Languages across boundaries: Studies in memory of Anna Siewierska*, D. Bakker and M. Haspelmath, Eds. Berlin/Boston: Walter de Gruyter GmBH, 2013, pp. 197–226.

[3] P. J. Hopper and E. C. Traugott, *Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press, 2003.

[4] T. Givón, "Topic, pronoun, and grammatical agreement," in *Subject and topic*, C. Li, Ed. New York, etc.: Academic Press, Inc, 1976, pp. 149–188.

[5] M. Ariel, "The development of person agreement markers: From pronouns to higher accessibility markers," in *Usage based Models of Language*, M. Barlow and S. Kemmer, Eds. Stanford: CSLI, 1999, pp. 197–260.

[6] D. Bickerton, *Roots of language*. Ann Arbor, Mich.: Karoma, 1981.

[7] M. A. Arbib, "Language evolution. An emergentist perspective," in *The Handbook of Language Emergence*, B. MacWhinney and W. OGrady, Eds. West Sussex, UK: Wiley/Blackwell, 2015, pp. 600–623.

[8] J. L. Bybee, *Morphology. A study of the relation between meaning and form*. Amsterdam/Philadelphia: John Benjamins, 1985.

[9] B. Heine and T. Kuteva, *The genesis of grammar. A reconstruction*. Oxford: Oxford University Press, 2007.

[10] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/

[11] M. Tomasello, *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press, 2003.

[12] R. Jackendoff, *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press, 2002.

[13] T. Givón, *Functionalism and grammar*. Amsterdam/Philadelphia: John Benjamins, 1995.

[14] S. Lestrade, "The space of case," Ph.D. dissertation, Radboud University Nijmegen, 2010.

[15] ——, "The optional use of morphological case," *Linguistic Discovery*, vol. 11, no. 1, 2013.

[16] P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: MIT, 2000.

[17] D. Nettle, *Linguistic Diversity*. New York: OUP, 1999.

[18] J. Bybee, *Language, Usage and Cognition*. New York: Cambridge University Press, 2010.

[19] J. R. Hurford, "Biological evolution of the saussurean sign as a component of the language acquisition device," *Lingua*, vol. 77, no. 2, pp. 187–222, 1989.

[20] E. Hutchins and B. Hazlehurst, "How to invent a lexicon: The development of shared symbols in interaction," in *Artificial societies: The computer simulation of social life*, N. Gilbert and R. Conte, Eds. London: UCL Press, 1995, p. 157189.

[21] L. Steels, "Constructing and sharing perceptual distinctions," *Machine Learning*, vol. ECML-97, p. 413, 1997.

[22] R. Van Valin, "Generalized semantic roles and the syntax-semantics interface," in *Empirical issues in formal syntax and semantics 2*, F. Corblin, C. Dobrovie-Sorin, and J.-M. Marandin, Eds. The Hague: Thesus, 1999, pp. 373–389.

[23] A. Siewierska, *Verbal Person Marking*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013.