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. 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.’
   c. text

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

(3) Standard Arabic [2, 10]
   a. ra’ay-tu 1-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 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

1The traditional term agreement cannot distinguish properly between the first two types, hence the new terminology.
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 anaphonic 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 seer (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 esodad. 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

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

³A 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.

⁴In 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.

⁵Both the dimensionality and distinctionality can be specified by the user, as virtually all other settings of the model.
least two dimensions).

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]).

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. [?]). 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 “pronomes”. 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
both need extra expression. An example is shown in (9):...
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

I would like to thank two anonymous reviewers for useful comments and the Netherlands Association for Scientific Research (NWO) for their financial support (grant 275-78-001, The exaptation of argument marking).

REFERENCES