Modeling Language Evolution: The emergence of argument-marking

Sander Lestrade

Centre for Language Studies, Radboud University Nijmegen, The Netherlands

Nothing in biology makes sense except in the light of evolution
(Theodosius Dobzhansky 1973)

What I cannot create, I do not understand
(Richard Feynman 1988)

1 Introduction

Language uncontroversially is the result of both *nature* and *nurture*: Something in our biological makeup enables us and only us to learn human language, but new generations of speakers at the very least will have to learn the cultural-specific pairings between words and concepts. A major question in linguistics has always been which parts of language go where; that is, which features are derived from primary data (and thus are "learned"), and which have to be activated only ("acquired"; cf. e.g. Plato's *Cratylus* or, more recently, Laurence and Margolis 2001; Gong, Shuai, and Comrie 2014). In the past decades, the learning camp seems to be gaining ground again: General learning mechanisms have been shown to be much stronger than originally expected (Rumelhart, Hinton, & Williams, 1986; Plunkett & Marchman, 1993), and cross-linguistic studies have revealed a wide variety of structures that are not easily parameterized into a single underlying language system that can be stored in the genome (Evans & Levinson, 2009). Moreover, recent insights in evolutionary cognition show that the brain massively reuses existing neural circuitry for simple tasks for new, cognitively more demanding processes (Anderson, 2010), and finally, it seems genetic change cannot keep pace with the rate of language change (Chater, Reali, & Christansen, 2009).

If rules of grammar indeed are to be derived from the input language by general learning mechanisms, this begs the question where the tendencies these rules capitalize on come from. To make things more concrete, this paper is concerned with the development of argument-marking constructions

n	noun marking	person indexing	word order
1	1	0	0
14	0	1	0
24	0	0	1
18	1	1	0
17	1	0	1
69	0	1	1
59	1	1	1

Table 1: **Frequency count of argument-marking systems as sets of strategies.** For the classification as present (1) or absent (0) of a strategy in a language, cf. Dryer (2013), Iggesen (2013), and Siewierska (2013). The sample here consists of the intersection of the languages studied by these authors. Most languages combine several strategies.

such as illustrated in (1). In this Turkish example, the verb indexes person features of its external argument (by means of -*um*) and the internal argument is marked for its role by a noun marker (-*i*). As the free translation shows, English uses word order for this instead, which completes the set of basic options languages use to make clear who does what to whom (Siewierska & Bakker, 2009).

(1) *Ben sen-i seviyor-um*
I you-ACC love.PROG-1SG
'I love you.' (Turkish; Derya Demircay, p.c.)

Again, the question is where such argument marking comes from. It is easily learned from the previous generation of speakers, perhaps, but this explanation does not work ultimately, as this would mean that the very first speakers had a full-fledged grammar already, as a result of which it would turn out to be innate after all. Alternatively, argument-marking strategies could have been invented at some point in time in order to improve communication. The problem with this proposal is that language does not seem to be the product of such intentional design. If this were the case, a single strategy would have sufficed, case marking arguably being most effective. Instead, as shown in Table 1 (taken from Lestrade, 2015c), language is overly complex (Premack, 1986; Bybee, Perkins, & Pagliuca, 1994; Gil, 2009): Most languages combine several strategies and especially case marking apparently is deemed insufficient on its own. Thus, it seems neither a simple learning nor a simple functional account will do. Is argument marking then innate after all?

The nature–nurture discussion recently gained depth by the distinction between *biological* vs *cultural evolution* (Deacon, 1997; Hopper, 1987; Saffran, 2001; Christiansen & Chater, 2008; K. Smith & Kirby, 2008; K. Smith, 2002; Chater et al., 2009). The first pertains to the biological makeup of living entities, works through genetic inheritance and is slow; the second concerns the artifacts creatures develop, works through learning, and is relatively fast. By recognizing the fact that the language system itself evolves too, the choice between learned and innate becomes less of a dichotomy, as innate (but general) learning dispositions mold language in the course of generations, making it an incredibly learnable system eventually. Assuming cultural evolution, we can save argument marking for nurture, as will be shown in this paper: We can now assume a first proto-stage of language that was qualitatively very different from modern language, with argument marking developing in the course of time only. Crucially, learning and development do not always go hand in hand in this process: Solutions that were motivated first may become generalized beyond communicative necessity eventually (cf. Durie's 1995 *functional overkill*; cf. also Dolan and Dayan, 2013), leading to such overly complex systems as shown in Table 1 indeed.

The relevant question under a cultural-evolution account now becomes *What are the principles and mechanisms that lead to the development of argument-marking strategies?* And an important assumption (or rather restriction) in answering this question will be that of *uniformitarianism* (originally a geologic notion by Charles Lyell; cf. a.o. Paul, 1891; Hopper, 1987; Croft, 2000; Saffran, 2001; Heine and Kuteva, 2007 for linguistic instances). According to uniformitarianism, the same principles and mechanisms that are currently operative have always operated in the past, and there is nothing else that molds language. Thus, the mechanisms and principles proposed should be observable in current language variation and change, and the emergence of argument marking should follow from extrapolation of these to prehistorical times. Note that this extrapolation works differently from the standard procedure in historical linguistics in which predecessor states are reconstructed from modern languages by backwards reasoning. Here, instead, an initial proto-stage of language is stipulated and successor stages are derived by going forwards in time.

1.1 Modeling language evolution

The problem for any evolutionary account of language is that language leaves no fossils.[1] We can only trace back the etymology of individual words 8,000 years or so (cf. Nichols, 1992, p. 313). Even if this were far enough to cover the origin of grammar, it would not say anything about it, as it does not tell us how grammar applied to these proto-words back then. At the same time, the oldest literary texts, which could have been informative in this respect, are much more recent and from "modern" languages with fully

[1] I don't know to whom to contribute this observation but it definitely is not mine.

developed grammars already. Hence neither do these say anything about the first emergence of argument-marking grammar. Fortunately, with modern computer technology, the fossil problem can be circumvented by *modeling* language evolution (Gong & Shuai, 2013; Cangelosi & Parisi, 2001; Jaeger et al., 2009).

In corroborating a cultural account using computational models, a fair procedure seems to be to attribute to culture those parts of language that can be explained by the application of more general principles, and accept language-specific machinery only for those parts that cannot be explained away (cf. the introductory remarks from above, and, for example, Jackendoff, 2002, p. 101, who argues that "[i]f some aspects of linguistic behavior can be predicted from more general considerations of the dynamics of communication in a community, rather than from the linguistic capacities of individual speakers, then they should be"). Following this line of argumentation, it has been shown that the development of compositionality and that of a conventional lexicon can be attributed to cultural evolution indeed (cf. Kirby, 2000; K. Smith, 2002; and Hurford, 1989; Hutchins and Hazlehurst, 1995; Jäger, 2007; Steels and Belpaeme, 2005; Vylder and Tuyls, 2006; Spranger, Pauw, and Loetzsch, 2010). Recently, also the cultural emergence of more complex language phenomena has been modeled successfully, such as that of agreement (Beuls & Steels, 2013; Lestrade, 2015b), morphological case (van Trijp, 2012; Lestrade, 2015a), spatial reference systems (Spranger et al., 2010), logical reasoning (Sierra-Santibáñez, 2015), irregular verbs (Pijpops, Beuls, & de Velde, 2015), aktionsart (Gerasymova, Spranger, & Beuls, 2012), and quantifiers (Pauw & Hilferty, 2012).

Unfortunately, however, these models generally are extremely limited in range and scope, because of which extrapolating their results to natural language phenomena is anything but trivial. Some simplification is of course necessary or even desirable when developing computational models (cf. for example Steels, 1998, p. 139, who argues that the goal is to "understand the principles, not mimick human language genesis in toto"), but as Galantucci (2005, p. 738) points out, there is a "wide gulf [. . .] between artificial agents and humans" because of which "drawing inferences from simulations to natural human phenomena is often problematic" (cf. Perfors, 2002; Jaeger et al., 2009; Lekvam, Gambäck, and Bungum, 2014 for similar concerns).

The model introduced in this paper, *MoLE*, for Modeling Language Evolution, tries to narrow the gap between artificial and natural language.[2] MoLE is distinctive for the degree of ecological validity that is pursued, both in terms of the motivation of the underlying principles and in terms of

[2]This paper introduces the current publicly available version of the model. Interim results (mostly of isolated developments) were reported in Lestrade (2015a, 2015b, 2016), in which the model still went under the working title *WDWTW*. Many things have changed in the course of the project and in case of contradictory statements, the present description should be considered conclusive.

the resemblance of the output with natural language. It assumes independently motivated, general communicative and cognitive principles only, and approaches argument-marking grammar holistically, with interacting rather than isolated strategies. As a result, the artificial languages that ultimately develop are equivalent to natural language, at least for present purposes.

An example of an artificial language that may emerge is exemplified in (2), in which content words are translated in real-world concepts for ease of understanding. As will be explained below in more detail, these translations are largely meaningless, as the agents know about abstract concepts only. What is important instead is the structure of the utterances, which is independent of the translation of the lexical items, of course.

(2) Lineage 1, 92^{nd} generation

a. *kymi ufisok-wi ahedybosi*
   2 throw-2 rock
   'You are throwing the rock.'

b. *iropiwu-od sumyhig fefogig-od*
   rabbit-U kangaroo see-3
   'As for the rabbit, the kangaroo is seeing it.'

The standard word order in this lineage after 92 generations is agent first, which is overruled in case of a contrastive topic, as shown in (2-b). Also, there is verb indexing for first and second person, as well as for preposed third-person topics (cf. *-wi* and *-od*). The topic marker *-od* is also used as a differential case marker in case of ambiguity (Bossong, 1985; P. de Swart, 2007): In the first example, it follows automatically who is throwing what given the propensities and affordances of the participants and the role requirements of the verb; but in the second, either the rabbit could see the kangaroo or the other way around (irrespective of topicality, cf. *It was the kangaroo that saw the rabbit*). To disambiguate, the marker *-od* is used to mark the less active participant or *undergoer* (the more active participant is called *actor*; terminology after Van Valin, 1999).

Importantly, these structures only develop over time. At the very beginning, agents made use of the *proto-language* illustrated in (3). That is, a language that consists of a conventional lexicon only, with lexical-ad hoc marking to resolve imminent communicative failure (cf. *okygokiri* in (3-b), which is used for the kickee role of the verb *kigefaryty* 'kick' only), and without verb agreement, standard word order, or any other form of (argument-marking) grammar. Hopefully the reader agrees that the utterances in (2) could be considered linguistic garden-variety, whereas very few modern languages follow the pattern illustrated in (3).

(3) lineage 1, proto-phase

a. *epimajuw adybasek odepykune*
grass cow eat
'The cow is eating grass.'
b. *adybasek ewinurib okygokiri kigefaryty*
cow horse kickee kick
'The cow is kicking the horse.'

In the next section, it will be explained how these results came about, and how the transition from (3) to (2) could have taken place. But first a final more general note on modeling language (evolution) is in place: According to A. D. Smith (2014, p. 282), there is an inherent tension in computational modeling between making a model more realistic and more analyzable. Whereas Smith argues that more complex models require more nuanced interpretation to understand, I think the opposite is true, in fact, at least with respect to the *implications* of the results. Any real-world complexity that is not implemented in a simple model needs to be taken into account in a proper discussion of its implications. This is easily done away with with some cursory statements in the discussion section, but should be taken much more seriously. The reason to use models in the first place is to study the interaction of factors that is beyond verbal argument (Niyogi, 2006). Thus, there is an inherent tension between understanding the technicalities and understanding the implications. This is not to say that simple models do not have any value; they very clearly do. But once the potential of a factor has been established in isolation, I think it is necessary to see how it behaves in interaction.[3]

2 MoLE

MoLE is available as an R package on CRAN (https://cran.r-project.org/). As just explained, as a consequence of pursuing ecological validity, MoLE is a complex model. Fortunately, its results can be understood and appreciated at a conceptual level too. Below, the model is discussed in two rounds, at a more general level first and in more technical terms next. Readers who want to study the codes in more detail are referred to GitHub (www.github.com/lestrade/MoLE) or the help pages of the package.

2.1 Background

Before discussing the model itself, it is important to make clear what it is and what it is not about, as its goals and results are easily misunderstood.

In MoLE, evolution is *not* about recombining parts of grammar to see which grammar eventually becomes dominant (according to a so-called *ge-*

[3]For what the comparison is worth, also in statistics, main effects can only be interpreted in light of their interactions.

netic algorithm). Agents are not born with a grammar, and the goal of the child is not to select a target grammar from a pre-specified hypothesis space (as e.g. in Niyogi, 2006). Also, the goal is not to provide time estimates for language change. It may be possible to tweak the model parameters such that the speed of a certain development corresponds to that in reality, but in principle, the number of generations it takes for some construction to emerge in the model is not meaningful. Finally, MoLE is not (yet) about population dynamics and social spread of a rule. The model easily allows for extensions in this direction, but at present these are not implemented.

(4) MoLE is *not* (yet) about
  a. choosing from an innate hypothesis space;
  b. recombining parts of grammar;
  c. providing time estimates for change, or
  d. population dynamics.

So what is MoLE about if not about the things in (4)? It's about identifying the necessary and sufficient general mechanisms and principles for grammar to emerge. Thus, it can be used to test cultural-evolution accounts of language by showing the feasibility of grammatical structures to emerge spontaneously. At present, it shows that argument marking can be modeled as the product of cultural evolution. But in fact argument marking should be considered a case study only. MoLE is developed in a modular way in order to allow for future extensions and test the evolution of other linguistic phenomena too (for which reason it is open-source). Again, the rationale of such experiments is that if it can be shown that linguistic structure develops spontaneously from more general functions, one need not assume language-specific cognitive machinery. The general goal, accordingly, is to show that rules of grammar are nothing but cultural conventions.

(5) MoLE is about
  a. identifying the underlying principles and mechanisms that lead to the development of grammatical structures;
  b. showing that grammar is nothing but a set of cultural conventions, and
  c. providing a model environment in which cultural-evolution hypotheses can be tested.

2.2 Conceptual idea

The general idea is that argument marking started out being communicatively motivated, as the result of *reflection*, in order to make sure the hearer correctly understood who does what to whom in an utterance. At some point, however, it became *reflexive*, i.e. automatized and generalized beyond communicative necessity (terms by Dolan and Dayan, 2013).

In the beginning of a simulation, agents are capable only of joint attention, intention-reading, and simple association, and, even more basically, of categorization and statistical learning (i.e., they are *language-ready*; Tomasello, 2003; Arbib, 2015). In addition, a desire for communicative success is stipulated, otherwise there is no need for them to make themselves understood. Crucially, agents do not have any form of grammar at this point, neither in terms of rules nor in terms of constructions. The only really language-specific feature agents have in the beginning is a conventional lexicon (which, as mentioned above already, can be shown to result from cultural evolution too).

Thus, as in other recent proposals about the cognitive architecture of language, the lexicon is put center stage (Vosse & Kempen, 2000; Jackendoff, 2002; Culicover & Jackendoff, 2005; Kempen, 2014). This lexicon is richer than traditionally understood, however, allowing for the life-time storage of the combinatorial information and usage history of all lexemes. It basically consists of object lexemes, or *nouns*, and action lexemes, or *verbs*. Both are pairings between a form and a set of meaning dimensions that collectively define the (referential) properties of the object or action they refer to (a la Gärdenfors, 2000; cf. also Katz and Fodor, 1963; Guiraud, 1968; Wierzbicka, 1996). In addition to describing actions proper, verbs also come with one or two verb-specific event-role descriptions. These role descriptions describe what the standard participants of the action have to be or do when the perspective of the particular verb is taken (cf. the (Neo-)Davidsonian approach in which an event argument is thought of as an argument itself, which needs to be characterized accordingly; Davidson, 2001; Parsons, 1994). It may seem redundant to specify both the action and the roles of its participants, but, on the one hand, an event involves more than just the activities of the core arguments (e.g., *cooking* involves heat and pans and is done for eating purposes, which does not follow from what the cook and the food themselves "do"), and also, it seems the very same event can be described using different perspectives and hence involves different argument roles (cf. *buy* vs *sell*, *eat* vs *eat a sandwich*, and *sweep the table* vs *sweep the crumbs*). The role whose perspective on the event is taken is called the *external*, the other one the *internal* role. For example, *sell* and *buy* basically describe the same event but whereas the former takes a source perspective, the second takes that of the goal. Note that this concerns the "lexical" perspective of the verb, not that of the overall utterance (cf. *John was sold the house*). Note further that these labels are for convenience only and do not have any cognitive status for the agents in the model.

It could also be hypothesized that utterances at this proto-stage consist of binary noun-verb combinations only (Progovac, 2015) or even that words refer to holistic events and do not make a distinction between action and object at all (Wray, 2000, cf. also Gil's 2005 concept of an *IMA language*, a maximally simple language that is morphologically isolating, syntactically monocategorial, and semantically associational). The combination into transitive constructions would then only start with the further development of a merge operation, or, in the second scenario, the differentiation between words for actions and objects would only start with the development of default role expectations, all event participants initially either requiring explicit marking or their contributions following by associative reasoning. Similarly, one could start from a phase in which duality of patterning did not develop yet, and hence there was no distinction between phonemes and morphemes (Hockett, 1982). Since such distinctions arguably developed before syntax did, otherwise, syntax would not have had anything to operate on, and since the goal is to study the development of grammar, these ingredients are simply taken for granted here. Note moreover that the assumption of an innate, or at least very early, distinction between actions and objects seems warranted given our predisposition to distinguish between these ontological types (cf. Bertenthal, 1996; Hurford, 2003).

Using Steels' 1997 *language-game* setup, two agents are sampled from the population and put in a situation in which multiple events are ongoing (cf. (6)). In each event, one or two events participants, possibly the speaker and/or hearer themselves, are involved in an action. The speaker has to find an adequate wording for one of the events, the *target event*, which is sufficiently distinctive given the other events in the situational context. The more similar the distractor objects and actions, the more specific the referential expressions have to be to single out their targets (Grice, 1975; Levelt, 1983).

(6) Situation

a. boy$_i$ is seeing man$_j$ (TARGET)
b. man$_j$ is reading book$_k$
c. boy$_{l/\not{i}}$ is eating
d. book$_k$ is red
e. woman$_m$ is walking dog$_n$
f. ...

In principle, the lexeme that best describes the intended referent will be selected to describe it. This is done by considering all meaning representations in the mental lexicon and determining their *semantic match*, that is, the degree to which the lexical representation of an item matches the properties of the referent. Additional factors that play a role in word selection are frequency and recency of use (which together constitute *word activation*), *collocation* and *collostruction* frequencies (combinations of words and words, and words and functions, respectively; Stefanowitsch and Gries, 2003), semantic weight (the more specific, the more difficult to retrieve), and economy of expression (the shorter the expression, the better). The word that scores best is considered first for expression.[4] On the basis of its semantic match with the distractor objects in the situational context (or events, for verbs), it is determined whether it is sufficiently distinctive to single out the target object indeed. If not, the second best word is considered until an expression is found that works. For example, given the presence of two different boys in (6), *boy* will not do to single out the one in the target event. Importantly, the lexeme that is eventually selected need not be the perfect expression, it only has to be sufficiently distinctive given the context. The expressions for the event ingredients that are thus selected together make up the basic proposition.

The order of the constituents in the utterance is determined by their activation initially. More activated (i.e. predictable, frequent and recent) elements are processed faster, and therefore ready for production earlier (Fenk-Oczlon, 1989; Levelt, 1983; Bock & Griffin, 2000; Schriefers, Teruel, & Meinshausen, 1998). Eventually, agents may observe certain tendencies in the ordering, on the basis of which they may form grammatical word-order generalizations, which will then overrule activation concerns.

Referring to events also involves making clear the distribution of event roles over the event participants in the communicated event. Oftentimes, this follows automatically, as for example in an eating event in which a man and a piece of meat are involved (7-a). If the distribution of roles does not follow automatically, however, the proposition has to be elaborated. Then, a prototypical performer of either of the roles is added as an ad-hoc role marker to make explicit the role of one of the participants (Aristar, 1997; Lestrade, 2010). For example, in an eating event with a man and a lion, *meat* can be used for disambiguation to make clear who is doing the "meat thing" in the eating event, as illustrated in (7-b).

(7) a. *meat man eat*
   'The man is eating the meat.'
   b. *man meat lion eat*
   'The lion is eating the man.'

In the model, markers are recognized as such by the hearer by the fact that an element would remain unanalyzed in the utterance otherwise. That is, an eating event normally involves two participants, the eater and the thing eaten, so if there are three additional lexemes uttered, one of these must be doing something else than refering. Using the general principle of *proximity* (Givón, 1995) and the stipulation that functional elements always follow referential expressions (Hawkins & Cutler, 1988; Hall, 1988),[5] the hearer

[4]For future versions of MoLE, it is of course desirable to substitute this serial selection procedure with a parallel one.

[5]Note that although neither assumption is strictly necessary, they conveniently restrict the hypothesis space for the hearer, which increases the speed of the simulation.

combines the vacant lexeme with the element preceding it, understanding it as a role marker. Note that it is not clear to the hearer a priori which of the elements the speaker is using referentially and which as role markers. That is, given a verb and three preceding elements as in (7-b), either the second could be the role marker of the first, or the third could be the role marker of the second. Alternatively, constituents can simply be ignored. The options for *man meat lion eat* with their corresponding interpretations are shown in (8):

(8) Possible analyses and interpretations of *man meat lion eat*

a. man=meat lion eat: 'The lion is eating the man.'
b. man meat=lion eat: 'The meat is eating the man.'
c. man_x ? ? eat: 'The man is eating.'
d. ? meat_x ? eat: 'The meat is eating.'
e. ? ? lion_x eat: 'The lion is eating.'
f. man_y ? ? eat: 'The man is eaten.'
g. ? meat_y ? eat: 'The meat is eaten.'
h. ? ? lion_y eat: 'The lion is eaten.'

All logical options are entertained by the hearer, and the one with the best match with any of the currently ongoing events is understood as the intended structure. This overall match is determined by establishing for each event in the situational context how well the (understood) actor participant in the utterance describes the actor of the event, and similarly for the undergoer participant and verb, with penalties for each word in the utterance that remained unanalyzed. The best matching event is now selected as the event the hearer thinks the speaker referred to. If the hearer thus correctly identifies the target event, the agents mark the successful usage of the words that constitute the utterance, remember the exact meaning for which the words were used, and next either switch turns to go on with their conversation or end it, after which two new agents are randomly selected for a new conversation.

During a conversation, agents build up a common ground of the things they have talked about. Objects in the common ground are more likely to be the topic of conversation by design, as they are more likely selected as ingredients in the events that constitute the context for the next communicative turn. Note that this involves meaning, not form, as the selection of the latter is dependent on the distractors in the context.

The communicative process is summarized in Figure 1. The speaker and hearer columns show more stable processes and components, the middle column the more fluid ones: The situation and utterance change completely with each communicative turn and the common ground with each conversation; instead, the lexicon and usage history are updated only, which has a much less profound impact.

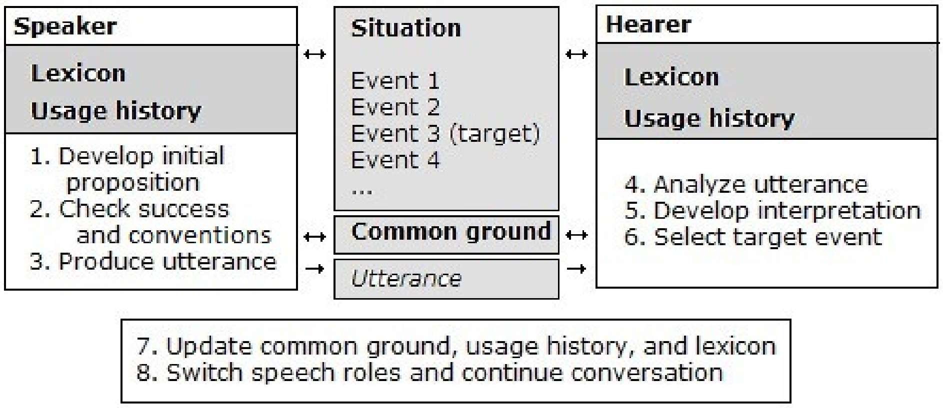

Figure 1: **Model representation of communication**

Some additional examples of utterances from a proto-phase of a different lineage are given in (9).[6] In (9-a), the role distribution should follow automatically and no extra role marking was deemed necessary; whereas (9-b) was judged ambiguous. Apparently, *fywydyjises* can be *umaremeled* just as well as *subydobyts* (or, in the made up translation, girls and boys are equally good students). Ad-hoc role markers can be used to resolve this ambiguity. As the glossing of *talibug* is meant to reflect, such markers are still very verb-specific at this stage. They are object terms themselves and the best performers of the verb role that currently is unclear. As with referential expressions, they need not be *perfect* performers: Here too, lexeme activation interferes and the first candidate that suffices to distinguish between the two roles (teacher from student, in this example) will be recruited.

(9) lineage 2, proto-phase

a. *odymirid subydobyt mihurenip*
read man book
'The man is reading the book.'

b. *fywydyjis umaremel subydobyt talibug*
girl teach boy student
'The girl is teaching the boy.'

As mentioned above already, agents do not keep talking like this. In the course of time, they develop grammatical structures and standard rules in a process of *grammaticalization* (Heine & Kuteva, 2007; Hopper & Traugott, 2003). Four principles play a main role in this process, viz. *generalization, fusion, erosion*, and *desemanticization*. Starting with the latter, words may extend their meaning range incidentally, either because the context does not require a more specific description than the most active one that presents

[6]Different lineages will be contrasted more systematically in Section 3.

itself first, 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 the word becomes more general. Whereas this arguably is an ongoing process in real life (Bybee, 2010), it is much easier implemented as an instantaneous event. Thus, shortly before they procreate, agents consider the variation in their usage history and may decide that certain meaning dimensions should either be changed or be removed from the meaning representation of a word. A meaning dimension of a word is changed into another value if this latter value is found significantly more representative for the meanings for which this word has been used. Deletion takes place if some meaning dimension is found to be unrepresentative or inconsistent with the uses of the word, but there is no dominant other value on that dimension among the uses.

In addition, frequent forms are more likely to be pronounced sloppily because of automatization. This is a production effect only and does not directly lead to different representations. If new generations of speakers have not yet firmly established their formal representation, however, these words may become represented accordingly by them eventually, which leads to erosion of form. (For, as Nettle, 1999 argues, sloppy pronunciation can only lead to shortening, not to elaboration of form.)

If a role marker becomes too short to stand on its own, it is suffixed to its head (Bybee, 1985). Thus, adpositions develop into suffixes (cf. also Hawkins and Cutler, 1988; Hall, 1988). The development of *pronouns* contains an additional stage. First, similar to role markers, if a referring expression becomes too short to stand on its own, it is suffixed to the verb. Unlike role markers, however, which only have to distinguish between two roles, referring expressions may become even too short to refer properly (Ariel, 1999). At this point, the pronoun, which already was suffixed to the verb, recruits another lexical item to take over its referential function. This item is preferably of the same person, but if these are not available, third-person lexemes may be recruited for this. Note that this process can for example be observed in modern spoken French, in which dative pronouns are used to help erst-while nominative pronouns that are clitized to the verb (e.g. *moi je=parle* 'me I=speak').

If combinations of referential expressions and role markers are very frequently used, they are no longer considered derived, but stored as wholes. This is especially likely for pronoun–role marker combinations, as individual lexical nouns generally do not appear frequently enough for this. Thus, more elaborate pronominal paradigms may emerge.

In sum, words develop more general meanings, because of which they can be used more widely and frequently, because of which their forms are more likely to erode and sometimes even fuse with with other words. After some time, pronouns such as *mepe*, verb markers such as *-se*, and role markers such as *-ta* may emerge, as shown in (10) from a speaker of the 27^{th} generation of lineage 2 (again, these numbers are not representative for developmental speed in natural language):

(10) lineage 2, 27^{th} generation

a. *mepe awijyf-se*
   1 go-1
   'I am going.'

b. *ydykumy-ta mofynodura=figa*
   cat-U chasing=dog
   'The dog is chasing the cat.'

The glossing of *-ta* in (10-b) as a generalized role marker, instead of the predicate-specific meaning *chasee*, follows from its generalized semantics and subsequent promiscuous use (to be explained in more detail in the next section). Unlike *talibug* in (9), which combines with *umaremel* 'teach' only and hence was glossed as 'student', *-ta* can be used for any undergoer role.

Note again that the initial use and development of the above structures and strategies can be understood as a solution to a communicative problem or the result of general processing mechanisms (to recapitulate, role markers are used when the role distribution does not follow automatically, verb indexing emerges when referential expressions lose the power to refer and need to recruit more expressive means for this, and word order follows from word activation). Their grammatical successors, however, are not always necessary for communicative success (Durie, 1995; Lestrade, van Bergen, & de Swart, 2016). At some point, these uses lose their communicative motivation and become *generalized* (a difference which according to Tanner, McLaughlin, Herschensohn, and Osterhout, 2013 even leads to different brain activation). Most likely, frequency plays an important role in this generalization: Speakers learn an association between a structure and the context in which it frequently appears (Bates & Elman, 1996; Dolan & Dayan, 2013; Croft, 2000), and if there are only few exceptions to the pattern observed, it is considered obligatory. In the more technical part below, it will be shown that the number of exceptions that is tolerated for a rule to develop in fact can be made explicit, using Yang's *Tolerance principle* (Yang, 2005, 2016).

Theories of language change differ in whom they assume responsible for such rule generalization, viz. children or adults. Croft (2000) argues that it cannot be children, for example because language change goes too slow, many acquisition errors are unlearned at a later age, it does not account for within-speaker variation, and children do not have the power to establish a change in conventions. Hudson Kam and Newport (2005), on the other hand, show experimentally that it is mostly children who regularize language, whereas adults faithfully copy the frequency distributions given in their experiments.[7] The present model is in fact ignorant in this matter:

[7] As an aside, I don't think the experiment of Hudson Kam and Newport (2005) pro-

Patterns can be observed and turned into rules throughout the life-time of the agents.

This concludes the conceptual tour. Hopefully, the intuition behind the development of argument-marking is now made clear. In the next section, it will be shown for a number of subroutines how they are implemented technically. Note again that it should be possible to appreciate the results in Section 3 on the basis of the above already, without reading through the technical details below. Thus, the reader may decide to skip the next part as a whole or check the operationalizations of interest only.

2.3 Technical implementation

In this section, the subroutines and design choices of the model that are most important for the assessment of the technical implementation are discussed, ordered by a combination of processing order and expected need for understanding. For a discussion of the other subroutines and/or further technical details, please refer to the help function of the R package or to the GitHub page.

Note that virtually all design choices are parameterized. If not mentioned otherwise, values discussed below are default settings only which can be set differently according to one's theory.

Meaning representation

Although agents in the model do not share our world, it is possible to implement natural-language concepts in a relatively valid way. According to Wierzbicka (1996), concepts can be decomposed into meaning primitives such as CONCRETE, HUMAN, MALE, etc. (cf. also Katz and Fodor, 1963; Guiraud, 1968). Similarly, in a way, Gärdenfors (2000) argues that concepts are sets of values on different meaning dimensions. Thus, we can think of a cat as something that is time-stable, concrete, alive, four-legged, tailed, etc. Note that whereas the initial meaning dimensions in such characterizations are very general and bisect the world (e.g. time-stability), additional meaning dimensions become more and more specific in order to single out a concept (e.g. having a tail). Abstracting away from the quality of the dimensions that organize our mental lexicon, the nominal lexicon of the agents are modeled as a list of randomly generated forms with values on a number of numerical meaning dimensions (their *vector representation*). Note that this abstraction is possible, since it is completely irrelevant what the vector

vides decisive evidence: The percentage of exceptions that needed to be "regularized away" exceeds the threshold for this as allowed by Yang's Tolerance principle (Yang, 2005, 2016). Moreover, Reali and Griffiths (2009) have shown that weak biases that are hardly observable in single generations may in fact lead to regularization in iterated learning (across generations, that is; cf. K. Smith, Kirby, and Brighton, 2003) representations mean to the agents; they only have to allow them to communicate about events and objects in their world, that is, to use referential expressions in the same conventional way as we do.

Following the observations just made, the default is for the dimensions to make an increasing number of distinctions, with values between 0–1. These dimensions may be taken to represent whatever properties are grammatically relevant for the linguistic behavior of words in natural language, but the model does not commit to any specific interpretation per se. Verbs are similarly specified (their dimensions simply representing different semantic features), with the additional vector specification of their predicate role(s). Examples of a noun and verb entry are given in (11) and (12).

(11) Meaning representation of nouns
wawelywul 1 0 0 1 1 0.75 0.5 0.5 0.625

(12) Meaning representation of verbs
nadesar 1 1 1 1 0 0.375 0.25 0.875 0.75
external role: 1 1 1 1 1 0.875 1 1 1
internal role: 0 0 1 0 0 0 0.125 0.375 0.375 0.125

Note that qualitatively, this is very different from the vector-semantics approach used in modern computational linguistics (e.g. Deerwester, Dumais, Furnas, Landauer, and Harshman, 1990; Mikolov, Chen, Corrado, and Dean, 2013), in which vectors represent behavior in texts rather than the underlying semantics that causes this behavior. Rather, the vectors used here should be understood as representations of activation in a neural-network model of the brain (Rumelhart, McClelland, & the PDP Research Group, 1986; Smolensky & Legendre, 2006). For example, the first dimension could be taken to represent a property that all animates objects do and inanimates do not have (i.e., it is activated in the vector representations of the first class only). Again, such specific interpretations are for explanatory purposes only, as dimensions do not have concrete meanings.

It should now be clear why, as was mentioned in Section 2.2 already, the glossing of the examples is largely meaningless. I will continue to provide concrete translations, however, as it makes it easier to see the relevant point, viz. the *structure* of the utterances.[8]

It is assumed in the model that new generations learn the lexicon from their parents without much problems. That is, they inherit the lexicon largely as is, with minor random modifications to those lexical items that have not been used by their parents at the point of procreation (but cf. *Usage history and meaning change* below).

[8]Thanks to Chara Tsoukala for this suggestion.

Prominence, agentivity, and actorhood

In the model, higher values, i.e. those closer to 1, are considered more prominent (for nouns) or agent-like (for roles) than lower values. Note that a correspondence is thus assumed between referential prominence and agentivity. That is, being prominent means being a good actor, and vice versa, and similarly for non-prominent undergoers.

As suggested by the verb representation in (12) already, if a verb has two role vectors, one of them contains higher values than the other on average. This is meant to reflect the semantic role distribution in natural languages, in which the predicate roles of two-place verbs generally can be distinguished in terms of agentivity, the external role being more agent-like than the internal one (Dowty, 1991; Van Valin, 1999; Yip, Maling, & Jackendoff, 1987; Primus, 1999).

At the beginning of a lineage, the two most prominent nouns are understood as having first and second person reference by default, all others are assigned a third person value. Thus, it follows automatically that local persons are prototypical agents (Silverstein, 1976). In later stages, as was explained in Section 2.2, other prominent items may acquire local reference too as a result of recruitment (Ariel, 1999; Ishiyama, 2012).

Lexeme selection 1: Vector comparison

The lexeme that best describes an intended referent will be selected to describe it in principle. This is determined by vector comparison of the referent with all meaning representations in the mental lexicon (using the `VMATCH` subroutine). First, the absolute difference between the target meaning and all lexical items on each meaning dimension D is determined. If lexical items are unspecified for certain meaning dimensions, the difference is considered zero. Next, these differences are weighed such that the first dimension is more important than the second, and similar for the second dimension with respect to the third, etc. (unless the `weigh` parameter is turned off). The rationale of this procedure is that in a taxonomy of meaning, the first few choices (e.g. concrete/abstract, animate/inanimate) are more crucially defining than lower ones (e.g. color). Technically, the differences on each difference are multiplied with the rank of their dimension, starting counting on the right. The *vmatch score* of a candidate c given a target t is 1 minus the average of these weighted differences, cf. (13). A toy example in which only four binary meaning dimensions are used is given in Table 2. In this competition, *umynefyb* describes the target meaning best.

(13) $$vmatch(c|t) = 1 - \frac{\sum_{j=i}^{n} |D_j^t - D_j^c| * weight_j}{\sum_{j=i}^{n} weight_j} + noise$$ with

(14) $$weight_j = n - j + 1$$

	meaning	abs. diffs	weighted	clean score	noised
target	1 0 0 0				
mykojada	1 1 1 0	0 1 1 0	0 3 2 0	0.5	0.509
umynefyb	1 0 1 0	0 0 1 0	0 0 2 0	0.8	0.799
pykamor	0 1 1 –	1 1 1 0	4 3 2 0	0.1	0.097
ofewatewe	1 1 0 1	0 1 0 1	0 3 0 1	0.6	0.604
idyjyhid	0 – 1 1	1 0 1 1	4 0 2 1	0.3	0.295

Table 2: **Word selection and vector comparison**. Meaning dimensions can be specified by 1 or 0, or remain unspecfied ("–"). In this competition, *umynefyb* describes the target meaning 1000 best.

Situation and event generation

For each communicative turn, a situational context is generated in which between 10 and 20 events are going on simultaneously. One of the events is the target event, which is the one the hearer should single out on the basis of the speaker's description. The other events provide the distractor objects and actions. The necessary degree of specificity for the expression differs with the number of distractor items.

Events are combinations of two or three vectors: one for the event action, one for the actor and, in case of a two-place event, one for the undergoer (cf. Table 3). The participants that figure in the situation are dependent on the *common ground* of the speech participants, which they develop while they talk. At the start of a conversation, it consists of the speech participants themselves and (the meanings of) three randomly selected lexical entries. With each turn, other objects have a small probability of entering the scene. Elements of the common ground may figure in multiple events at the same time (for example, I am presently both writing this text and breathing). After Du Bois (1987), the odds for a new Actor argument are 1/30 and those for a new Undergoer 1/4.

On the basis of these objects' propensities and affordances, action predicates are sampled from the mental lexicon. For this, the vector match is determined between the properties of the object and all verb roles corresponding to the assigned generalized role of the object (i.e., whether it is acting as an Actor or Undergoer). These *typing* scores are then used as the sampling probabilities for the corresponding events (terminology after Aristar, 1997).[9] The impact of the typing scores, and hence the match between object qualification and role requirement, is set by the parameter `roleNoise`. By default, roughly one third of the verbs will never be combined with a given object and most are only incidentally so; only the top 5% of best-matching verbs has a reasonable probability of being combined

[9]Thanks to Sebastian Collin for helpful suggestions in the design of this procedure.

V1	...	V9	A1	...	A9	U1	...	U9	target
0.00	...	0.12	0.00	...	0.50				0
0.00	...	0.12	0.00	...	0.25				0
1.00	...	0.75	1.00	...	1.00	1.00	...	0.38	1
1.00	...	1.00	0.00	...	1.00	1.00	...	0.38	0

Table 3: **First entries of a situation (abbreviated).** *V1:V9* show the properties of the actions, *A1:A9* the referential properties of the actors, *U1:U9* the ones of the undergoers, the target column identifies the event of interest. Situations are consistuted by between 10 and 20 events.

with a particular object.

For each two-place predicate that is thus selected, a second argument is sampled from the common ground or mental lexicon (using the same odds as above) on the basis of its role-qualifications for the vacant role (using again typing scores, but now the other way around, from role to performer).

Table 3 shows an (abbreviated) example of a situation in the model. The *V* columns refer to the characteristics of the actions that are ongoing, *A* refers to the referential properties of the actors, and *U* to those of the undergoers (if present). The *target* column identifies the event that is to be communicated. Which exact semantic roles the participants get depends on the particular verb that is chosen by the speaker to conceptualize the event (e.g. as walking, running, or striding).

Note that through this setup events in fact consist of instantiated mental meaning representations (with the addition of some random noise). This is of course not very realistic, as the mind does not determine what happens in the real world. There, things work exactly the other way around, the concepts we have being generalizations of the things we perceive. But we can exploit this connection the other way around with the same net result, which seems highly desirable for ecological validity: Both in reality and in the model, there is considerable overlap in what people/agents think and what they experience. That is, the probability to find a dog barking is higher than for it to read (and the other way around for a student). Everything being parameterized, however, the virtual world can be made entirely random by setting `roleNoise` and `refNoise` to 1.

Usage history and meaning change

For each lexeme it is remembered how it was used. Thus, if *umynefyb* were used to express the meaning 1000 indeed, as in Table 2, this usage would be added to the usage history. Role markers, which do not have a referential use, are stored with the vector representation of the roles they mark. Crucially, these uses may differ from the lexical representation of the items: The “real” meaning of *umynefyb* is 1010, whereas it now has been used for 1000.

Shortly before two agents procreate and endow their offspring with their lexicon (by mixing their lexicons, each providing half of the items), they consider the variation in their usage history. For each meaning dimension, it is considered if the uses are in line with the lexical representation. If not, two things can happen: If a certain meaning value clearly stands out (in terms of the Tolerance principle, cf. the discussion of *Generalization* below), the lexical representation is adjusted accordingly. For example, if *umynefyb* were almost always used for things with a zero value on their third dimension, the original one is substituted by a zero.

If there is no dominant meaning value to be found, however, the meaning dimension with the greatest variation is emptied, as a result of which the lexeme becomes underspecified for this dimension. For example, if *umynefyb* were used for things with a zero value on their third dimension only half of the time, whereas the other dimensions are more often in line with the meaning representation, the specification of the third dimension is removed from the representation of *umynefyb* altogether.

For this latter option, a frequency threshold is assumed that increases (non-linearly) with the number of dimensions that have been removed already. For a first dimension to be removed, an item has to be used at least in .8% of the utterances, the second dimension requires 1.6%, the third 4.0%, up to 40% of all uses for the prefinal dimension to be removed. These values are determined heuristically, the intuition being that it should be relatively easy to become moderately specified, but it takes extreme frequencies to become extremely general.

Lexeme selection 2: Candidate ordering

The most important factor in lexeme selection has already been discussed, viz. the semantic match with the referent (or, for role markers, with the role specification). In addition, *resting activation* plays a role too, as frequently and recently used words are more "activated" and present themselves before less activated words (Balota & Chumbley, 1985; Fenk-Oczlon, 1989; Levelt, 1983; Bock & Griffin, 2000). Resting activation is implemented by means of the formula in (15). The main intuition is that activation is frequency divided by recency. The initial addition of 1 to frequency is necessary to allow for the log transformation of unused items; a small `recencyDamper` is added to recency in order to dampen its effect (without it, the activation of the second most recent item would only be half of that of the most recent one, which seems too drastic a decay). A bit of noise is added to the result, and, finally, the result is rescaled to the 0–1 range by the `RESCALE` function to make it compatible with the semantic match score.

(15) $$activation(lexeme_i) = RESCALE(\frac{log(frequency_i+1)}{recency_i+recencyDamper} + noise)$$

Other factors in word selection are linguistic context (in the form of *collostruction* frequencies, i.e. combinations of words and functions; Stefanowitsch and Gries, 2003), semantic weight (the more specific, the more difficult to retrieve), and economy of expression (the less constituting characters, the better). Collectively, these factors determine the `candidateScore` of an item as specified in (16)

(16) $$\begin{aligned} &candidateScore(lexeme_i) = \\ &\quad semantic\ match_i + \\ &\quad activationImpact * resting\ activation_i + \\ &\quad collostructionImpact * RESCALE(collostruction_i) + \\ &\quad semWeightImpact * min(semWeight)/semWeight_i + \\ &\quad economyImpact * min(productionEffort)/productionEffort_i \end{aligned}$$

with default (relative) impact values of .2 for resting activation and collostruction frequency, and .1 for semantic weight and economy of form. As can be seen, collostruction frequency is rescaled to the 0–1 range, and both semantic weight and production effort are relative to the best option available (because of which they fall in the same range). Because of this more complex procedure, a lexeme with a perfect meaning match can be outranked by a more frequent or recently used item that is not too far off.

The candidate with the highest score is considered for expression first. On the basis of its semantic match with all distractor objects in the situational context (or distractor events, for verbs), it is determined whether it is sufficiently distinctive to single out the target. For this, the match with the target has to be .05 higher than that with any of the distractors (given a range of 0–1).

Erosion and suffix development

Words are pronounced sloppily by the speaker if they have been used frequently (in more than 5% of the utterances) or recently (less than two utterances ago), or are predictable because of the textual context (used in a particular collostruction more than thrice). Sloppy pronunciation involves the simple deletion of the final character of a form, for example using *umynefy* instead of *umynefyb*. If the hearer has used this lexeme less than three times (or rather: if it has used the lexeme it thinks corresponds to the form produced by the speaker less than three times, cf. *Form recognition and utterance decomposition*), its form representation is changed accordingly.

If the word length falls below 4 characters, it can no longer be used independently as a result of which it is suffixed to its head. This concerns the preceding host argument for role markers and the verb for referential expressions.

Check proposition and elaborate

The speaker constructs its initial proposition by selecting sufficiently distinctive lexemes, as explained in *Word selection and vector comparison.* Before actually producing an utterance, speakers determine whether an utterance can be formulated by simply producing the forms of the ingredients in order of activation (through the `CHECKSUCCESS` procedure). There are three reasons for elaboration.

First, according to Ariel (1999), words can no longer refer properly once they become too attenuated. In the model, the `referenceThreshold` is set to 2 characters. If a referential expression is 2 characters or less, it is suffixed to the verb, after which a new, longer word is selected for expression, which results in an agreement relation between the attenuated suffix and supporting element. Since local-person expressions erode most strongly, this mostly applies to local pronouns. If there are no corresponding local-person expressions available, a third-person item is recruited for this. In this case, upon recognition of the retired pronoun, the speaker of course has to overwrite the person value of the recruited item, after having identified which of the referential expressions in fact was the recruited one. In case of doubt, the most prominent expression is considered the newly recruited one.

The second reason for elaboration is to make sure the distribution of roles is understood correctly. Speakers check whether the typing score of an argument for its intended roles is distinctively higher than that of the other participant and than its score for the other role (i.e., under the reverse distribution of roles). If so, the hearer should be able to derive the meaning. If not, a marker is added to make the role distribution explicit. The default for this is to use the best marker available for either of the roles. The speaker checks how the external role can be expressed and how the internal one can, and then determines which of the two is to be preferred (given their relative semantic matches and usage frequencies; see *Lexeme selection 2: Candidate ordering*). Alternative options are for example to mark the worst performer, the second argument, or either of them randomly.

The third reason for elaboration, and in fact the one that is considered first as it may preempt the others, is that a generalization has been made that applies in the current situation (cf. *Generalization*). For example, a convention may exist to put Actor arguments in first positions (because of which they are recognizable even with bad typing scores) or to always index the person of the Actor participant on the verb.

Form recognition and utterance decomposition

`FMATCH` compares a perceived form with the forms of the lexeme entries stored in the lexicon to search for the best matching entry. This match is determined using editing distance, in which deviations towards the onset are considered worse than deviations at the offset of words, as the former are generally pronounced more carefully than the latter (following the *cohort model* of word activation; cf. Hawkins and Cutler, 1988, p. 295 and references therein). Thus, *gabodur* is considered a better match to *gaboduf* than to *sabodur*.

A hearer explores all possible decompositions of the words in an utterance by trying if any of the suffixes in its lexicon (i.e., words with forms with less than 4 characters) can be aligned word finally (taking into consideration possible reduction in pronunciation). Thus, if an agent knows the word *gabod*, it entertains the possibility that the speaker says *gabo*, which is less than 4 characters long, because of which it would be used as a suffix. If it now hears the word *yfetegabo*, it will try to find entries for both *yfetegabo* as a whole and *yfete* and *gabo(d)*, in which the latter is used as a marker on *yfete*.

Verb identification and constituent analysis

To crack an utterance, a hearer first has to identify the verb. It searches for a form match for each item in the verbal lexicon. Next, it considers the best action match in the situation for each verb candidate. The candidate with the best overall match (combining form and meaning) is identified as the verb. Once the verb has thus been identified, all remaining forms are matched up with entries from the nominal lexicon. Next, all combinatorial possibilities are considered for constituents, as was illustrated by the *man meat lion eat* example in (8).

Pronoun development

Forms may erode into suffixes, as explained above in *Erosion and suffix development*. Thus, role markers can develop into maximally short case markers. For referential expressions this is not possible. If they fall below the referential threshold two four characters (Ariel, 1999), they can no longer refer properly. Since erosion applies to frequent items mostly, and local expressions are used most frequently, it is mostly local expressions that fail to meet the threshold. If a speaker notices an expression falls short, it selects a supporting element to establish the referential relation. Since there is only one expression per local person in the beginning, this involves the recruitment of a prominent third-person item initially. If this third-person item is frequently recruited for supporting a local reference, its person features may become overwritten accordingly, at which point a new local pronoun is established.

The challenge for the hearer during the transition period, in which the new person value has not been established yet, is to recognize the local-person recruit as such. In the example in (17), both *egomo* 'man' and *jokywypu* 'goat' in principle have third-person values. Because of the verb marker *-yw* the hearer knows there is a first-person participant involved. And since the hearer knows *-yw* cannot refer itself because it's too attenuated, one of the remaining elements must have been recruited for this (alternatively, *-yw* is ignored in the analysis and the first-person reference is missed altogether). The element that is most prominent and has most often been used in this role is selected as the recruit. The analysis in (17-a) is an example from an actual simulation, the one in (17-b) only shows the unlikely logical alternative.

(17) Lineage 5, 4^{th} generation

a. *egomo jokywypu balamisub-yw*
man/1 goat see-1
'I'm seeing the goat.'

b. *egomo jokywypu balamisub-yw*
#man goat/1 see-1
#'I'm seeing the man.'

Fusion

In the model, fusion of referential expressions with role markers involves a recombination of their expressions to create a new form. This form combines with the vector specification of the original referential expression in which the meaning dimensions for which the marker is (still) specified overwrite the corresponding values of the new referential expression. The new item is added to the lexicon; the constituting ingredients can still be used independently.

In the example below, the second-person pronoun *pepesu* is fused with the undergoer marker *ep*. The constituting characters are mixed and the prominent meaning values of *pepesu* are overwritten with the undergoer values of *ep*. This results in a pronominal paradigm for the second person of, what now has become, the Actor pronoun *pepesu* and the new Undergoer pronoun *sesupupe* (cf. *I* vs *me*).

(18) Lineage 4, 28th generation

a. pronoun: *pepesu* 1 1 1 1 1 0.875 – – –
b. marker: *ep* 0 – 0 0 0 – – – –
c. fusion: *sesupupe* 0 1 0 0 0 0.875 – – –

For two items to fuse, they have to be used in more than 2% of the utterances, which is a tougher criterion than it may seem, as local persons only appear as Undergoers in 3% of the utterances (according to the `dahl0` parameter; following Dahl, 2000) and it need not be the Undergoer role that gets marked for disambiguation moreover (given `bestMarker` as `solutionMethod`, it may also be the Actor participant that is marked for its role; cf. *Check proposition and elaborate*).

Generalization

At some point, languages develop conventional ways of doing things by applying solutions even in the absence of their motivating problems (Durie, 1995; Bresnan, Dingare, & Manning, 2001; Lestrade et al., 2016). For example, instead of marking only those prominent Undergoers that cannot be recognized as such, simply all prominent Undergoers may become marked. Such automatization can be hypothesized to save effort: The costs for always using an Undergoer marker for a prominent performer without thinking may be less than for checking all prominent Undergoers only to find out it is necessary to mark them again and again (Lestrade et al., 2016).

Yang (2005, 2016) proposes there is even a mathematically determinable threshold for a rule to emerge (and be maintained). If the number of times a potential rule is found not to apply is smaller than this threshold, it can become a rule indeed. This threshold θ is determined by the formula in (19), which says that the maximum number of exceptions to a rule is the number of times it applies in principle, divided by the natural logarithm of that number.[10]

(19) $\theta_n = \frac{n}{ln(n)}$

Thus, in our example, if the number of prominent Undergoers that follow automatically in spite of their prominence is below the total number of prominent Undergoers divided by their log, all prominent undergoers will be marked; if too many prominent Undergoers can do without marking, however, no general rule will develop.

The same reasoning can be applied to the development of person indexing and grammatical word order, of course: If the number of agents that is not found in first position is below the threshold, a general rule will develop that says to put all agents in first position, and if the number of times a local participant is not indexed on the verb is below the threshold, all local participants will be indexed, no matter their degree of attenuation.

3 Results

In this section, the results of the simulations using various parameter settings will be discussed. All examples discussed above are from a "world" with default parameter settings, but many other worlds are possible. For

[10] I do not agree with Yang's psycho-linguistic explanation for this threshold, the discussion of which is beyond present purposes; I do accept, and conveniently use, his empirical observation, however.

world	description (non-default parameter setting)
default	as discussed in Section 2.3
noCheck	no check for communicative success (`checkSuccess=F`)
binary	only binary meaning distinctions (`distinctions=rep(2,9)`)
lexical	only non-binary distinctions (`distinctions=c(rep(9,9))`)
match	candidate ordering determined by semantic match only (`candidateOrder='match'`)
frequency	candidate ordering determined by frequency only (`candidateOrder='frequency'`)
collostruction	increased impact of collustruction frequency on candidate ordering (`collostructionImpact=.3, frequencyImpact=.1`)
noLinking	no preference for prominent external roles (`linkingPreference=1`)
talkAge	newborn agents join conversation straight away (`talkAge=0`)

Table 4: Simulated worlds

example, agents may not check whether or not they are understood, and may use binary meaning distinctions only (instead of using increasing distinctionality). For a number of world types, seven lineages each are run to check for within-consistency. The assumption is that if something happens six out of seven times, it shows a genuine preference.[11] The general hypothesis is that the development of argument marking is independent from these settings. Whenever a difference in outcome is observed, this should be explained in terms of the difference in world parameters accordingly. The worlds, listed in Table 4 involve what seem to me the most speculative, or controversial, modeling assumptions. Given 64 world parameters, there are of course many other worlds possible, which the reader may wish to explore for themselves.

In Lestrade (2016), it was shown that nothing changes to the language system if words are not allowed to erode or extend their meaning. Since this is trivial, such a world is not simulated here.

Below, different argument-marking strategies will be discussed in turn, first showing their development in *default* lineages, and next discussing their development in the other worlds, explaining the differences and similarities. But first, let us consider the development of communicative success in the course of the generations.

3.1 Communicative success

As shown in Figure 2, all lineages are quite successful from the start, and maintain their success throughout the simulation. Whatever happens in the course of time to distinguish between the worlds then, it is not so much

[11]This is very loosely after Fisher's tea test, which requires 5/5 positive results. In language, however, nothing is without exceptions, hence the 6/7 requirement.

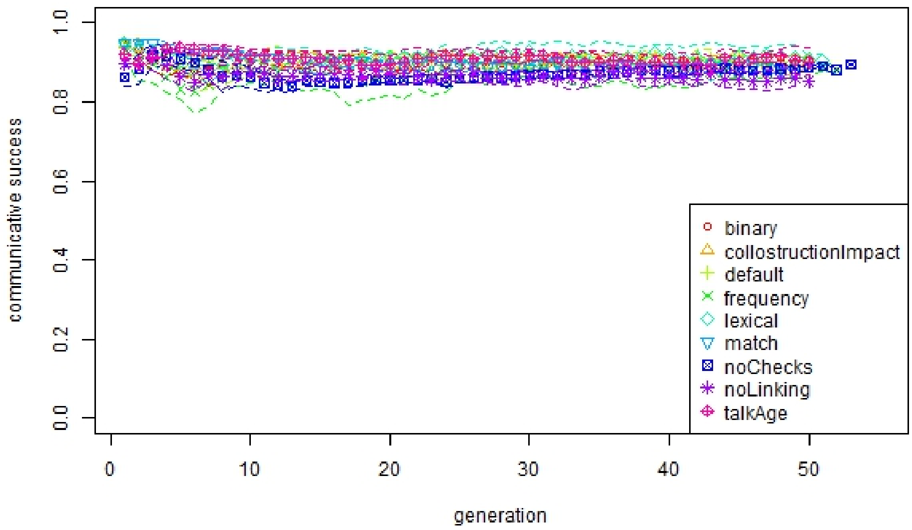

Figure 2: **Communicative success per world type.** Each world type is tested with seven lineages. Dashed lines show standard deviations.

in terms of communicative success. Simulations are run equally long, but the *noChecks* lineages take less computation power because of which they develop slightly more generations.

Although all lineages are highly successful, lineages from the *frequency, noLinking*, and *noChecks* world do seem to score a bit lower consistently (Wilcoxon rank sum test Ws 3.950, 2.378, and 3.891, respectively with all ps approaching zero, rendering multiple-comparison correction unnecessary; cf. Figure 3 in the appendix for single contrasts). These results can easily be understood. If *frequency* is the only factor in candidate selection, there is only a minimal test for overall meaning correspondence (cf. *Check proposition and elaborate*). Slight deviations in meaning representations now become more easily fatal for communicative success. If roles are not clearly distinct, as in the *noLinking* world, it is harder to keep them apart. And finally, if nothing is done to prevent miscommunication whenever an intended meaning does not follow automatically (*noCheck*), obviously the meaning will not follow automatically.

Nevertheless, even in these worlds communicative success is around 90% and the differences in success, although significant, can hardly be considered relevant. This follows logically from the predictability of the world, which is quite high and kept constant for all worlds. Moreover, given only 20 different events that constitute a situation, it may suffice to reconstruct only a single participant from the utterance to correctly identify the corresponding event. In sum, given a sufficiently predictable world and a limited context, it is really not that hard to communicate successfully.

3.2 Case marking

Case marking (or *role marking*, *noun marking*, or *flagging*) is sometimes considered a simple binary feature of language, languages being described as either having it or not. With the recognition of *differential* case-marking systems (Bossong, 1985; P. de Swart, 2007; de Hoop & de Swart, 2008), the classification becomes more nuanced already, but arguably, the picture is even more complex, differential case marking being either *local* or *global* (P. de Swart, 2007, 2011; Malchukov & de Swart, 2009). A global DCM system uses case marking only if the distribution of verb roles were unclear otherwise, as it cannot be told from the context or lexical semantics of the ingredients (e.g. given two equal participants: *boy sees man*). A local DCM system uses marking for certain role-performer combinations irrespective of whether this leads to a presently unpredictable meaning (e.g. all inanimate Actors: *wind breaks glass*).

In the model, four stages of case marking can be discerned in principle, the fourth and ultimate stage never being reached (although it is attested in natural language). First, in very early phases, role marking is entirely ad hoc and lexical, which could be considered the absence of case marking really. Next, generalized role markers may develop which are used when necessary only, if communication failed otherwise. This phase corresponds to global DCM. Two types of (not necessarily mutually exclusive) generalization may develop from this phase. Lineages may develop default marking for local persons through the development of pronominal paradigms (see Section 3.3 below) or they may decide to use case marking if performers have certain particularly atypical meaning values. Both of these initially led to frequent case marking in a global system and now have been generalized into a default, thus establishing a local system. Finally, lineages may decide to use case marking for a generalized role by default, irrespective of the type of performer. In natural language, this seems to be quite rare as most languages seem to be of the differential type (Sinnemäki, 2014). An example could be Hungarian, although this really represents a fifth stage in the development of case marking that is not implemented in the model (yet), viz. the *structural use* of case. For case to become used structurally, it is necessary that languages develop *grammatical roles* first, i.e. combinations of generalized roles and discourse functions (Comrie, 1981). Coupling these in the model is still left for future work, however.

This scarcity of the fourth (and fifth) stage of case marking, both in the model and cross-linguistically, is not unexpected as the number of exceptions to such a rule is much too large. Indeed, in a predictable world, the cases in which role marking is necessary should be the exception. This suggests that, at least in some languages, case marking is not only used for disambiguation and additional motivations play a role in its development. An obvious candidate that is often proposed is the *identification function* (Mallinson & Blake, 1981; de Hoop & Malchukov, 2008), whose grammaticalization arguably follows from the same reasoning (Lestrade, 2010). The implementation of this function too is for future research.

Table 5 in the Appendix shows for each lineage what kind of role markers were used eventually. Role markers are listed if they are used more often as a role marker than as an argument and if they have lost at least two meaning dimensions. A role marker is considered a *case* marker if it eroded into a suffix and lost at least four out of nine meaning dimensions (i.e., if it's maximally short and maximally general; Lestrade, 2010).

3.2.1 Case development in default world

With the exception of the fifth lineage, all lineages of the *default* world developed case marking eventually. In the fifth lineage, the competition between the two Undergoer markers *-uh* and *-wo* has not been settled yet, because of which neither is frequent enough to develop into a case marker. Case markers, if they develop, generally have extreme role score on their binary meaning dimensions, having lost all their non-binary distinctions. These extreme scores can be explained by the preference for (natural-language) predicate roles to differ in prominence (Dowty, 1991; Van Valin, 1999; Yip et al., 1987; Primus, 1999): If a verb has two role vectors, one of them contains higher values than the other on average. As a result, a prototypical marker can be expected to have corresponding meaning values, marking either prototypical Actors or Undergoers.

Marking either of these generalized roles should be sufficient, as the other role follows automatically. Thus, it can be expected that only one marker develops per lineage. With the exception of the third lineage, in which both an Actor and Undergoer marker are used, this holds true indeed.

For third person arguments, case marking remains global in all lineages of the default world. For role marking of local persons, cf. Section 3.3.

3.2.2 Case development in other worlds

Role markers can be identified in most lineages of the other worlds. Some lineages did not develop role markers, however, e.g. the fourth and seventh lineage of the *collostruction Impact* world. This can be explained by the interaction with other argument-marking strategies, as will be shown in Section 3.6. In the *noChecks* worlds, none of the lineages developed role markers whatsoever. This follows logically, as speakers in these lineages do not bother to make the distribution of roles explicit in the first place. checken

Case markers do not develop as easily as role markers by definition, as they have to grammaticalize further. In some worlds they never do, in fact, viz. *binary, match*, and *noLinking* (and, obviously, *noChecks*). Note that if the role markers of a lineage do not meet the case criteria, the most frequent ones mostly have extreme role scores still.

Non-binary meaning dimensions are more easily lost than binary ones, as specific uses are more likely to be slightly off: Half of the potential uses of an item corresponds to either a zero or one on its binary meaning dimensions, but the more fine-grained the distinctions, the more often the actual uses differ from the lexical representations. Thus, it can be understood why markers develop into case markers in the *binary* world less easily. Also the lack of case markers in the *match* lineages follows logically. Candidate selection (for role markers) is based on meaning match only, because of which items are less often used outside of their comfort zone, because of which they generalize less easily. Finally, if there is no linking preference, as in the *noLinking* lineages, generalizations are less easily made too. Differently from in the other worlds, predicate roles do not group into a distinct prominent and non-prominent class. Whatever grouping is made within the roles eventually, an underlying (cognitive) motivation is lacking, because of which generalized Actor and Undergoer roles do not apply. As a result, more, and less extreme markers are called for, with a more uniform frequency distribution.

Differently from the *default* lineages, the (case) markers that develop in the *lexical* world, in which all underlying meaning dimensions are multi-valued, all have intermediate role scores. I'm not sure whether this should be considered a technical quirk or a relevant finding (but cf. the discussion section). In any event, the difference follows logically: Whereas the linking preference holds equally for binary and non-binary dimensions, the outcome is rather different. Binary meaning dimensions have a simple 5:1 distribution (given the default 5:1 setting of the linking parameter, with five ones for each zero on prominent role dimensions, and the other way around for the non-prominent ones). But these odds are distributed over the values of non-binary dimensions instead: The one most extreme value indeed is five times more likely than the other extreme, but this difference develops only gradually over the in-between distinctions. Since lexical meaning specifications of multi-valued dimensions are more likely to be (sligthly) off, as just explained, and since no single value is dominant enough to characterize the role-marker uses, multi-valued dimensions are more likely to be deleted than to be rewritten. But whereas the markers in the other worlds thus remained their binary specifications (which, again, were likely to be all zeros or ones), the meaning dimensions that remain in the lexical world are more nuanced.

No local-DCM systems developed distinguishing within third-person participants in any of the worlds. Since this did happen in previous simulations in which no interaction between argument-marking strategies was allowed (e.g. Coupe and Lestrade, to appear), this seems to be due to the interaction with word order (cf. below). Local DCM did develop for person, however, in the form of a pronominal case paradigm for first and second person.

3.3 Pronominal paradigms

HIER

3.3.1 Pronominal paradigms in default world

Paradigm: Pronominal paradigm is recognized as such if variation in pronoun is due to fusion of pronoun with marker.

Paradigms short-lived.

Local Undergoer pronouns short-lived: Actor pronoun likely to develop into verb suffix, after which Undergoer pronoun takes over Actor role and adapts maning accordingly.

Development into verb-suffix should be more difficult, for example blocked by mere presence of another suffix.

3.3.2 Pronominal paradigms in other worlds

3.4 Verb marking

agreement, indexing

3.4.1 Verb marking in default world

3.4.2 Verb marking in other worlds

3.5 Word order

Hammarstrom JoLE 2016: in terms of families with a preference for a certain word order, SOV is much more common than anything else (with 56,6%, SVO being prefered by 13% of the families). Families are more interesting to count, as the preferences of related languages are not independent (but at this level, SOV 43,3%, SVO 40,3%).

Siewierska (1988): Basic word order in "stylistically neutral, independent, indicative clauses with full noun phrase participants, where the subject is de

nite, agentive, and human, the object is a de

nite semantic patient, and the verb represents an action, not a state or an event."

Lehmann Fundamental Principle of Placement (FPP) voor. Het gaat alleen om de primary concomitants V en O, S is minder sterk verbonden en soms zelfs optioneel. Resultaat: OV vs VO.

Venneman denkt ook dat de bevindingen van Greenberg vooral om V en O gaan, maar gebruikt die anders. Het Principle of Natural Se- rialization (PNS) zet alle operatoren (i.e. dependents/ modi

ers) en operands (heads) in dezelfde volgorde.

Hawkins: twee typen taal geeft: prepositioneel en postpositioneel. Hij stelt het Principle of Cross-Category Harmony (PCCH) voor (cf. Song 22): The more similar the position of operands relative to their operators across different operand categories considered pairwise (verb in relation to adposition order, noun in relation to adposition order, verb in relation to noun order), the greater the percentage numbers of exemplifying languages.

Tomlin zelf stelt de volgende principes voor (5) a. Theme First Principle (TFP): thematische argumenten (oude informatie) komen voor nietthematische (nieuwe informatie; en S and Theme correleren) b. Animated First Principle (AFP): animater eerst, en als dat niet werkt gebruik je de semantische rol: Agent X Instrumental X Benefactive/ Dative X Patient (en de meest animate NP wordt S) c. Verb-Object Bonding Principle (VOB) volgt uit wat we eerder al zeiden en Behaghels First Law: wat semantisch (qua betekenis) bij elkaar hoort komt syntactisch (qua vorm) ook bij elkaar (ook wel: ico- niciteitsprincipe, Haiman))

Kiss 225-6: in WALS, correlation between SOV/SVO and finite/nonfinite averbial and relative clauses. Via Hawkins: minimal constituent recognition domain if inf-marker of embedded clause (SOV) or subordinate marker (SVO) close to verb.

Primus (1999, Case and Thematic roles): A before P is simplest, because P presupposes A. Kemmerer D (2012, "The crosslinguistic prevalence of SOV and SVO word orders...), "A instigating action" event hardwired in brain

Word order speaker or hearer driven? Buy speakers and hearers time to process phonologically, semantically and/or syntactically complex, new or heavy constituents (Arnold et al., 2000; Wasow, 2002). There has been some discussion about whether word order is primarily affected by complexity, length or newness and it also does not seem to be clear how to define these notions exactly in the first place, cf. Tily (2010). Bates and MacWhinney (1987, 223): "points will be understood better if their topic is understood in advance", but here that is a lucky result at best. Or: Constituents can partly be predicted by the preceding words, and therefore the sentence-initial position is least predictable (Ferrer-i-Cancho 2014). The topic is the most predictable unit in a sentence, independent of its relation to the other constituents. Overall predictability improved by putting topic first.

3.5.1 Word order in default world

3.5.2 Word order in other worlds

3.6 Interactions

Trade-off between strategies?

4 Discussion

As for the intermediate role scores of the lexical world. I'm inclined to think that this does (yet) not show a deeper cognitive truth. First try different technical procedure: not just right/wrong and delete dimension with most variation, but take into account distance between lexical specification and mean use.

Different explanations possible. Per uniformitarianism, same factors that drive modern preference should have been active in first stages, so then too SOV. Other type of explanation (Newmeyer 2000, Gell-Mann and Rhulen 2011, Maurits and Griffiths 2014): merely coincedence. Proto-language happende to be SOV (for no special and certainly reason), hence most modern descendants are. Those who are not, drifted away.

First reconstruction: Lehman 1973, 1974. Barddal (2015, routledge handbook): mostly critized for presenting typological tendencies as universals

More generally, it is shown here that the development of an important part of grammar can be explained in terms of *cultural evolution*, meaning that it is learned rather than innate Deacon, 1997; Saffran, 2001; Hopper, 1987; Croft, 2000; Heine and Kuteva, 2007; K. Smith and Kirby, 2008. Introduction argued against innate, but it still needs to be shown how grammatical phenomena could develop in practice, of course, which is exactly what the present paper tried to do.

Relative impact of various factors that determine word selection. We know a word's frequency/resting activation plays a role (resting activation is the brain's equivalent of frequency). Also, we know that semantic priming plays a role (you are faster to say *cat* after hearing *dog* than after hearing *book*, as does constructional priming (there are only so many things that can follow *I enjoyed reading that ...*). And obviously, meaning itself is a factor: if you want to refer to a house, you consider words that match this concept. Although the relative impacts among various sorts of priming and resting activation could be experimentally determined, there is (as of yet) no simple way to determine semantic match in real life. It cannot be expressed straightforwardly in a single number.

Minor issues: given two verb suffixes, the internal one is understood as refering to the internal argument/"object" of the verb. Problematic as, first, except for this, external vs internal arguments do not play a role whatsoever in the model, no status. (Initially, that is; one could make the agents generalize over these roles and have them develop a notion of lexical subject). Second, although cross-linguistic preference for internal-argument verbal suffix to be closer to the external-argument one indeed, this is not without counter-examples: DGfS.

Unresolved: if choice between simple and morphological complex word, how to choose which analysis is right (without calculating all possibilities?

Special focus: word order (through activation)

5 Conclusion

In sum, case markers straightforwardly develop as a result of grammaticalization. As relative frequency plays a role in word activation, lexical items that have previously been used for role disambiguation are more likely to be considered again. As there are maximally only two roles to be kept apart, role markers often need not be very specific. Thus, a previously used marker is often found good enough, as a result of which its frequency of usage increases further, as well as the variation of its usage contexts. Because of the former, its form is likely to erode; because of the latter, its meaning is likely to bleach. Eventually, this may lead to a the model equivalent of a case marker (Lestrade, 2010): a maximally short form with a minimal number of meaning dimensions specified that marks its host for its function or type of dependence.

Although the synchronic functionality of indexing is debated, its development too can easily be modeled. Recall that speech participants are part of the common ground by default and therefore figure in many events. Because of the resulting frequent and varied usage, words referring to local persons are prone to erosion and desemanticization. But differently from role markers, which do not have a referential function, once the form of a referential expression becomes too short to refer properly, a more expressive copy has to be added. As a result, erstwhile local pronouns end up indexing the person of their helpers, and verb-indexing mostly develops for local persons only. Word-order generalizations, finally, also follow straightforwardly. Again, local person are often talked about as they are standard part of the common ground. Importantly, they are prototypical agents too. As a result, agents are more activated on average and therefore tend to appear first in an utterance. Also the order of verb and undergoer follows from such general considerations: Since the verb is less often topical and more abstract, its average activation is lower than that of the undergoer which therefore generally comes before it (Schouwstra & de Swart, 2014; Hoeks, 2016). As a result of both effects, a tendency for Actor-Undergoer-Verb (or "SOV") order may emerge. A competing generalization that may interfere in this development is to put topics in first position (which are also more activated on average and have to be put in first position if used contrastively moreover).

Acknowledgments

This research was made possible by the Netherlands Organization for Scientific Research (NWO; Veni grant 27578001, *The exaptation of argument marking*), whose financial support is gratefully acknowledged. Thanks to Chara Tsoukala and Sebastian Collin for helpful suggestions.

References

Anderson, M. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*, 245–313.

Arbib, M. A. (2015). Language evolution. An emergentist perspective. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 600–623). West Sussex, UK: Wiley/Blackwell.

Ariel, M. (1999). The development of person agreement markers: from pronouns to higher accessibility markers. In M. Barlow & S. Kemmer (Eds.), *Usage based models of language* (pp. 197–260). Stanford: CSLI.

Aristar, A. R. (1997). Marking and hierarchy. Types and the grammaticalization of case markers. *Studies in Language*, *21*(2), 313–368.

Balota, D. A. & Chumbley, J. I. (1985). The locus of word-frequency in the pronunciation task: Lexical access and/or production? *Journal of memory and languages*, *24*, 89–106.

Bates, E. & Elman, J. (1996). Learning rediscovered. *Science*, *274*, 1849–1850.

Bertenthal, B. I. (1996). Origins and early development of perception, action, and representation. *Annual Review of Psychology*, *471*, 431–459.

Beuls, K. & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLOS ONE*, *8*(3).

Bock, K. & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning. *Journal of experimental psychology: General*, *129*(2), 177–192.

Bossong, G. (1985). *Empirische Universalienforschung. Differentielle Objektmarkierung in der neuiranischen Sprachen.* Tübingen: Narr.

Bresnan, J., Dingare, S., & Manning, C. (2001). Soft constraints mirror hard constraints: Voice and person in english and lummi. In M. Butt & T. Holloway King (Eds.), *Proceedings of the LFG 01 conference.* CSLI Publications.

Bybee, J. (1985). *Morphology. a study of the relation between meaning and form.* Amsterdam/Philadelphia: John Benjamins.

Bybee, J. (2010). *Language, usage and cognition.* New York: Cambridge University Press.

Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar. Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press.

Cangelosi, A. & Parisi, D. (2001). Computer simulation: A new scientific approach to the study of language evolution. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language.* London: Springer.

Chater, N., Reali, F., & Christansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *PNAS*, *106*(4), 1015–1020.

Christiansen, M. H. & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–509.

Comrie, B. (1981). *Language universals & linguistic typology* (2nd). Chicago: University of Chicago press.

Coupe, A. R. & Lestrade, S. (to appear). Non-structural case marking in tibeto-burman and artificial languages. To appear in *Linguistic Discovery.*

Croft, W. (2000). *Explaining language change: An evolutionary approach.* Harlow etc.: Longman.

Culicover, P. & Jackendoff, R. (2005). *Simpler syntax.* Oxford, etc.: Oxford University Press.

Dahl, Ö. (2000). Egophoricity in discourse and syntax. *Functions of Language*, *7*(1), 37–77.

Davidson, D. (2001). The logical form of action sentences (1967). In *Essays on actions and events* (pp. 105–122). Oxford: Clarendon Press.

Deacon, T. (1997). *The symbolic species.* London: Penquin.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Dolan, R. & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619.

Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, *63*(4), 805–855.

Durie, M. (1995). Towards an understanding of linguistic evolution and the notion 'x has a function y'. In W. Abraham, T. Givon, & S. A. Thompson (Eds.), *Discourse grammar and typology grammar and typology: papers in honor of john w.m. verhaar* (pp. 275–308).

Evans, N. & Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and brain sciences*, *32*, 429–49. doi:doi:10.1017/S0140525X0999094X

Fenk-Oczlon, G. (1989). Word frequency and word order in freezes. *Linguistics*, *27*, 517–556.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*, 737–767.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought.* Cambridge, MA: MIT.

Gerasymova, K., Spranger, M., & Beuls, K. (2012). A language strategy for aspect: Encoding aktionsarten through morphology. In L. Steels (Ed.), *Experiments in cultural language evolution* (pp. 257–276). Amsterdam: John Benjamins.

Gil, D. (2005). Isolating-monocategorial-associational language. In H. Cohen & C. Lefebvre (Eds.), *Categorization in cognitive science.* Oxford: Elsevier.

Gil, D. (2009). How much grammar does it take to sail a boat? In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 19–33). Oxford: Oxford University Press.

Givón, T. (1995). *Functionalism and grammar.* Amsterdam/Philadelphia: John Benjamins.

Gong, T. & Shuai, L. (2013). Computer simulation as a scientific approach in evolutionary linguistics. *Language Sciences*, *40*, 12–23.

Gong, T., Shuai, L., & Comrie, B. (2014). Evolutionary linguistics: theory of language in an interdisciplinary space. *Language Sciences*, *41*, 243–253.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). New York: Academic Press.

Guiraud, P. (1968). The semic matrices of meaning. *Social Science Information*, *7*(2), 131–139.

Hall, C. J. (1988). Integrating diachronic and processing principles in explaining the suffixing preference. In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 321–349). Oxford, UK: Basil Blackwell.

Hawkins, J. A. & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 280–317). Oxford, UK: Basil Blackwell.

Heine, B. & Kuteva, T. (2007). *The genesis of grammar. a reconstruction.* Oxford: Oxford University Press.

Hockett, C. (1982). The origin of speech. In W. S.-Y. Wang (Ed.), *Human communication: Language and its psychobiological bases* (pp. 4–12). Reprint of Scientific American (1960) 203, 88-111.

Hoeks, M. (2016). From SOV towards SVO. Explaining the word order distribution in terms of changing preferences. BA thesis, Radboud University Nijmegen.

de Hoop, H. & de Swart, P. (Eds.). (2008). *Differential subject marking.* Dordrecht: Springer.

de Hoop, H. & Malchukov, A. [Andrej]. (2008). Case-marking strategies. *Linguistic Inquiry*, *39*(4), 565–587.

Hopper, P. J. (1987). Emergent grammar. In *Proceedings of BLS* (Vol. 13, pp. 139–157).

Hopper, P. J. & Traugott, E. C. (2003). *Grammaticalization* (2nd). Cambridge: Cambridge University Press.

Hudson Kam, C. & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*, *1*(2), 151–195.

Hurford, J. R. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, *77*(2), 187–222.

Hurford, J. R. (2003). The neural basis of predicate-argument structure. *Behavioral and brain sciences*, *26*, 261–316.

Hutchins, E. & Hazlehurst, B. (1995). How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial societies: The computer simulation of social life* (pp. 157–189). London: UCL Press.

Iggesen, O. A. (2013). Number of cases. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Ishiyama, O. (2012). The diachronic relationship between demonstratives and first/second person pronouns. *Journal of Historical Pragmatics*, *13*(1), 50–71.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jaeger, H., Steels, L., Baronchelli, A., Briscoe, T., Christiansen, M. H., Griffiths, T., ... Triesch, J. (2009). What can mathematical, computational and robotic models tell us about the origins of syntax? In D. Bickerton & E. Szathmáry (Eds.), *Biological foundations and origin of syntax* (Vol. 3, pp. 385–410). Cambridge, MA: MIT Press.

Jäger, G. (2007). Evolutionary game theory and typology: A case study. *Language*, *83*(1), 74–109.

Katz, J. & Fodor, J. (1963). The structure of a semantic theory. *Language*, *39*(2), 170–210.

Kempen, G. (2014). Prolegomena to a neurocomputational architecture for human grammatical encoding and decoding. *Neuroinform*, *12*, 111–142.

Kirby, S. (2000). Syntax without natural selection. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language* (pp. 303–323). Cambridge, UK.

Laurence, S. & Margolis, E. (2001). The poverty of the stimulus argument. *The British Journal for the Philosophy of Science*, *52*(2), 217–276.

Lekvam, T., Gambäck, B., & Bungum, L. (2014). Agent-based modeling of language evolution. In *Proceedings of 5th workshop on cognitive aspects* *of computational language learning (cogacll)* (pp. 49–54). Gothenburg: EACL.

Lestrade, S. (2010). *The space of case* (Doctoral dissertation, Radboud University Nijmegen).

Lestrade, S. (2015a). A case of cultural evolution: The emergence of morphological case. *Linguistics in the Netherlands*, *32*, 105–115.

Lestrade, S. (2015b). Simulating the development of bound person marking. In H. Baayen, G. Jäger, M. Källner, J. Wahle, & T. Baayen-Oudshoorn (Eds.), *Proceedings of the 6th conference on quantitative investigations in theoretical linguistics*. Tuebingen: University of Tuebingen.

Lestrade, S. (2015c). The interaction of argument-marking strategies. In S. Lestrade, P. de Swart, & L. Hogeweg (Eds.), *Addenda. artikelen voor ad foolen* (pp. 251–256). Nijmegen: Radboud University.

Lestrade, S. (2016). The emergence of argument marking. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference* (Vol. 11).

Lestrade, S., van Bergen, G., & de Swart, P. (2016). On the origin of constraints. In L. et al. (Ed.), *Optimality-theoretic syntax, semantics, pragmatics*. Oxford: Oxford University Press.

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Malchukov, A. [A.] & de Swart, P. (2009). Differential case marking and actancy variations. In A. Malchukov & A. Spencer (Eds.), *The Oxford handbook of case* (pp. 339–355). Oxford: Oxford University Press.

Mallinson, G. & Blake, B. J. (1981). *Language typology: Cross-cultural studies in syntax*. Amsterdam: North-Holland.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781 [cs.CL].

Nettle, D. (1999). *Linguistic diversity*. New York: OUP.

Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago/London: The University of Chicago Press.

Niyogi, P. (2006). *The computational nature of language learning and evolution*. Cambridge MA: The MIT press.

Parsons, T. (1994). *Events in the semantics of English: A study in subatomic semantics*. Cambridge, MA: MIT press.

Paul, H. (1891). *Principles of the history of language*. Translation of *Principien der Sprachgeschichte*, by H.A. Strong. London: Longmans, Green.

Pauw, S. & Hilferty, J. (2012). The emergence of quantifiers. In L. Steels (Ed.), *Experiments in cultural language evolution* (pp. 277–304). Amsterdam: John Benjamins.

Perfors, A. (2002). Simulated evolution of language: A review of the field. *Journal of Artificial Societies and Social Simulation*, *5*(2).

Pijpops, D., Beuls, K., & de Velde, F. V. (2015). The rise of the verbal weak in inflection in Germanic: An agent-based model. *Computational Linguistics in the Netherlands Journal*, *5*, 81–102.

Plunkett, K. & Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, *48*(1), 21–69.

Premack, D. (1986). *Gavagai! or the future history of the animal language controversy*. Cambridge, Mass.: MIT Press.

Primus, B. (1999). *Cases and thematic roles. Ergative, accusative and active.* Tübingen: Max Niemeyer Verlag.

Progovac, L. (2015). *Evolutionary syntax*. Oxford, UK: Oxford University Press.

Reali, F. & Griffiths, T. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*, 317–328.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(9), 533–536.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, Mass.: MIT Press.

Saffran, J. (2001). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114.

Schouwstra, M. & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*, 431–436.

Schriefers, H., Teruel, E., & Meinshausen, R. M. (1998). Producing simple sentences: Results from picture-word interference experiments. *Journal of memory and language*, *39*, 609–632.

Sierra-Santibáñez, J. (2015). An agent-based model of the emergence and transmission of a language system for the expression of logical combinations. In *Proceedings of the twenty-ninth aaai conference on artificial intelligence*.

Siewierska, A. (2013). Verbal person marking. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Siewierska, A. & Bakker, D. (2009). Case and alternative strategies: word order and agreement marking. In A. Malchukov & A. Spencer (Eds.), *The Oxford handbook of case* (pp. 290–303). Oxford: Oxford University Press.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Atlantic Highlands: Humanities Press.

Sinnemäki, K. (2014). A typological perspective on differential object marking. *Linguistics*, *52*(2), 281–314.

Smith, A. D. (2014). Models of language evolution and change. *WIREs Cogn Sci*, *5*, 281–293. doi:10.1002/wcs.1285

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*, *14*(1), 65–84.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial life*, *9*(4), 371–386.

Smith, K. & Kirby, S. (2008). Cultural evolution: Implications for understanding the human language faculty and its evolution. *Phil. Trans. R. Soc. B*, *363*, 3591–3603.

Smolensky, P. & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, Mass.: MIT Press.

Spranger, M., Pauw, S., & Loetzsch, M. (2010). Open-ended semantics co-evolving with spatial language. In *The evolution of language (evolang)* (Vol. 8).

Steels, L. (1997). Constructing and sharing perceptual distinctions. *Machine Learning, ECML-97*, 4–13.

Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, *103*, 133–156.

Steels, L. & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, *28*(4), 469–89.

Stefanowitsch, A. & Gries, S. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209–243.

de Swart, P. (2007). *Cross-linguistic variation in object marking* (Doctoral dissertation, Radboud University Nijmegen).

de Swart, P. (2011). Sense and simplicity: Bidirectionality in differential case marking. In A. Benz & J. Mattausch (Eds.), *Bidirectional Optimality Theory* (pp. 125–149). Amsterdam/Philadelhpia: John Benjamins.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of l2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367–382.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

van Trijp, R. (2012). The evolution of case systems for marking event structure. In L. Steels (Ed.), *Experiments in cultural language evolution* (pp. 169–205). Amsterdam: John Benjamins.

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

Vosse, T. & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.

Vylder, B. D. & Tuyls, K. (2006). *Journal of Theoretical Biology*, *242*, 818–831.

Wierzbicka, A. (1996). *Semantics: primes and universals: primes and universals*. Oxford University Press.

Wray, A. (2000). Holistic utterances in protolanguage. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language* (pp. 285–302). Cambridge, UK: Cambridge University Press.

Yang, C. (2005). On productivity. In *Linguistic variation yearbook* (Vol. 5, pp. 265–302).

Yang, C. (2016). *The price of linguistic productivity. How children break the rules of language*. Cambridge, MA: MIT Press.

Yip, M., Maling, J., & Jackendoff, R. (1987). Case in tiers. *Language*, *63*, 217–250.

Appendix

Communicative success

Role-marking paradigms

The *marker* column shows the form, the *freq(uency)* column the total frequency of use as a role marker (in 2000 utterances). The weight column shows the number of meaning dimensions that are still specified, the score column the average value on the these dimensions. Recall that values towards zero correspond to Undergoer meaning, values towards one to Actor meaning. Markers are listed only if they have lost at least two meaning dimensions. A marker is considerd a *case marker* if it eroded into a suffix and lost at least four out of 9 meaning dimensions (cf. the final column). If a lineage is absent, no role markers developed in it (e.g. the fourth lineage of the *collustructionImpact* world).

Table 5: **Role-marking paradigms**

world	lineage	marker	freq.	weight	score	case
default	1	-*opyw*	147	5/9	.8	1
		-*tu*	142	6/9	0	0
		-*oj*	48	7/9	.3	0
	2	-*ud*	452	5/9	0	1
		-*ru*	26	7/9	0.3	0
	3	-*pu*	328	5/9	1	1
		-*ag*	211	5/9	0	1
		-*el*	6	7/9	.6	0
	4	-*fo*	287	5/9	0	1
		-*ef*	165	6/9	0.9	0
		-*yk*	19	7/9	0.2	0
	5	-*uh*	67	7/9	0.2	0
		-*wo*	63	7/9	0.2	0
	6	-*am*	306	5/9	0	1
		-*it*	68	7/9	0.8	0
		-*fo*	23	7/9	0.4	0
		-*os*	11	7/9	0.4	0
	7	-*yb*	268	5/9	0	1
		-*jy*	258	6/9	1	0
		-*or*	7	7/9	0.1	0
binary	1	-*ar*	204	6/9	0	0
	2	-*ri*	354	6/9	0	0
	3	-*mi*	277	7/9	1	0
		-*eh*	229	6/9	0	0
		-*uh*	49	7/9	.9	0

Table continued on next page.

Table 5 continued from previous page.

world	lineage	marker	freq.	weight	score	case
	4	-*if*	219	7/9	1	0
		-*lo*	208	6/9	0	0
	5	-*ly*	270	6/9	0	0
		-*id*	219	7/9	1	0
binary	6	-*lb*	423	6/9	0	0
		-*oh*	82	7/9	1	0
		-*us*	16	7/9	.1	0
	7	-*et*	433	6/9	0	0
		-*im*	53	7/9	.3	0
	1	-*ow*	203	6/9	0.1	0
		-*ih*	46	7/9	0.3	0
	2	-*lu*	136	6/9	0	0
		-*ob*	60	6/9	.9	0
		-*je*	33	7/9	.7	0
		-*oh*	28	7/9	1	0
collostructionImpact	3	-*ji*	208	5/9	1	1
		-*an*	53	6/9	0	0
		-*ma*	16	7/9	.3	0
	5	-*fe*	69	6/9	0	0
	6	-*yr*	188	5/9	0	1
		-*hi*	67	6/9	0.2	0
	1	-*bo*	305	5/9	0	1
		-*mu*	286	5/9	1	1
		-*ew*	31	7/9	.8	0
	2	-*li*	395	4/9	1	1
		-*at*	314	5/9	0	1
		-*ib*	1	5/9	1	1
	3	-*ek*	377	5/9	1	1
		-*me*	369	6/9	0	0
		-*de*	1	7/9	.4	0
	4	-*ha*	361	6/9	.1	0
		-*py*	327	6/9	1	0
frequency	5	-*yg*	338	5/9	1	1
		-*ge*	321	5/9	0	1
	6	-*ym*	318	5/9	1	1
		-*ei*	351	5/9	0	1
	7	-*ip*	312	6/9	1	0
		-*ri*	159	7/9	.2	0
		-*yg*	74	7/9	.9	0
		-*yk*	68	6/9	.2	0
		-*ol*	23	5/9	1	1
		-*ud*	18	6/9	.5	0
		-*ul*	7	7/9	.4	0
		-*fu*	5	6/9	.6	0

Table continued on next page.

Table 5 continued from previous page.

world	lineage	marker	freq.	weight	score	case
lexical	1	-*ib*	323	5/9	.6	1
		-*aw*	105	6/9	.5	0
		-*jy*	38	7/9	.2	0
		-*yw*	25	7/9	.6	0
		-*uj*	8	7/9	.5	0
	2	-*ep*	27	6/9	.5	0
	3	-*ba*	245	5/9	.2	1
		-*ji*	165	6/9	.4	0
		-*as*	152	6/9	.4	0
		-*ih*	76	7/9	.2	0
		-*ur*	2	7/9	.4	0
	4	-*od*	18	7/9	.3	0
	5	-*uh*	1135	3/9	.4	1
		-*yg*	108	6/9	.4	0
		-*ed*	9	6/9	.5	0
		-*on*	67	7/9	.4	0
		-*ym*	13	7/9	.4	0
		-*hu*	4	7/9	.6	0
		-*yg*	1	7/9	.4	0
	6	-*ok*	395	5/9	.5	1
match	1	-*ig*	67	7/9	.1	0
		-*no*	48	7/9	.1	0
		-*hy*	42	6/9	.2	0
		-*yt*	26	7/9	.9	0
		-*ob*	19	7/9	.1	0
		-*as*	17	7/9	.8	0
	2	-*ly*	44	6/9	1	0
		-*ew*	39	7/9	0	0
		-*ip*	39	7/9	.1	0
		-*el*	39	7/9	.4	0
		-*si*	25	7/9	0	0
		-*ob*	19	7/9	.9	0
		-*ug*	16	7/9	.1	0
		-*te*	14	7/9	.1	0
	3	-*wa*	17	7/9	.9	0
	4	-*ow*	62	7/9	.9	0
		-*bo*	37	7/9	0	0
		-*el*	35	7/9	.6	0
		-*pi*	34	7/9	.1	0
		-*ek*	30	7/9	.4	0
		-*fy*	30	7/9	0	0
		-*uf*	24	7/9	.2	0
		-*yn*	19	7/9	1	0

Table continued on next page.

Table 5 continued from previous page.

world	lineage	marker	freq.	weight	score	case
match	5	-*ak*	35	7/9	.1	0
		-*wig*	32	7/9	.1	0
		-*ih*	29	7/9	.1	0
		-*ob*	28	7/9	.1	0
		-*yd*	23	7/9	0	0
		-*of*	19	7/9	.9	0
		-*ly*	16	7/9	.2	0
	6	-*uj*	58	7/9	.3	0
		-*ty*	50	7/9	.9	0
		-*ji*	36	7/9	.1	0
		-*ga*	27	7/9	.1	0
		-*ab*	24	7/9	.2	0
		-*dy*	17	7/9	.1	0
	7	-*ni*	94	7/9	.9	0
		-*al*	33	7/9	.1	0
		-*eb*	28	7/9	.9	0
		-*ho*	27	7/9	.1	0
		-*yp*	18	7/9	0	0
		-*eh*	13	7/9	.9	0
		-*no*	11	7/9	.6	0
noLinking	1	-*ku*	157	7/9	.4	0
		-*ek*	96	6/9	.8	0
		-*be*	56	6/9	.7	0
		-*ly*	48	6/9	.1	0
		-*il*	14	7/9	.7	0
	2	-*ew*	124	6/9	.6	0
		-*he*	72	7/9	.6	0
		-*ko*	59	6/9	.4	0
		-*ryg*	55	7/9	.3	0
		-*ru*	44	6/9	.3	0
		-*ra*	2	7/9	.5	0
	3	-*sy*	55	7/9	.5	0
		-*du*	50	7/9	.6	0
		-*ka*	42	7/9	.8	0
		-*gy*	40	7/9	.7	0
		-*uk*	27	7/9	.3	0
		-*ed*	23	7/9	.7	0
		-*ih*	17	7/9	.3	0
		-*do*	5	7/9	.8	0
	4	-*or*	70	7/9	.3	0
		-*gu*	45	7/9	.7	0
		-*so*	32	7/9	.8	0
		-*up*	32	7/9	.4	0
		-*em*	20	7/9	.4	0
		-*ew*	15	7/9	.1	0

Table continued on next page.

Table 5 continued from previous page.

world	lineage	marker	freq.	weight	score	case
noLinking	5	*-od*	173	6/9	.9	0
		-yt	78	7/9	.2	0
		-sa	31	7/9	.4	0
		-if	25	7/9	.4	0
	6	*-us*	102	7/9	.9	0
		-im	60	7/9	.4	0
		-ti	24	7/9	.3	0
	7	*-ot*	126	6/9	.4	0
		-sy	79	7/9	.7	0
		-it	76	6/9	.5	0
		-by	52	7/9	.3	0
		-dy	3	7/9	.3	0
talkAge	5	*-me*	433	5/9	0	1
		-yf	78	7/9	.3	0
		-to	15	7/9	.4	0
	6	*-bi*	61	5/9	0	1
	7	*-oj*	8	7/9	.1	0

Local case-marking generalizations

Generations at which local case-marking generalizations are made for person. If lineages (or worlds) are absent, no generalization were made. *Generalization* column shows which participants are marked (and at which generation, if differentiated).

Table 6: **Local person marking**

world	lineage	generation	generalization
default	2	50	2U
binary	2	50	1U
	3	9, 10	2U
	5	22, 23	1U
	6	19, 20, 27, 28, 43	1U (20, 27, 28), 2U (19, 20, 43)
collostruction	6	31	1U
frequeny	2	41	1U
	4	32:34	1U (33), 2U (32, 33), 3U (32:34)
lexical	5	26:50	1U (26:30,), 2U (27:39, 43, 44, 49, 50), 3U (27:50)
match	2	47	1U
	3	27, 46	1U (27), 2U (46)
	7	42	1U

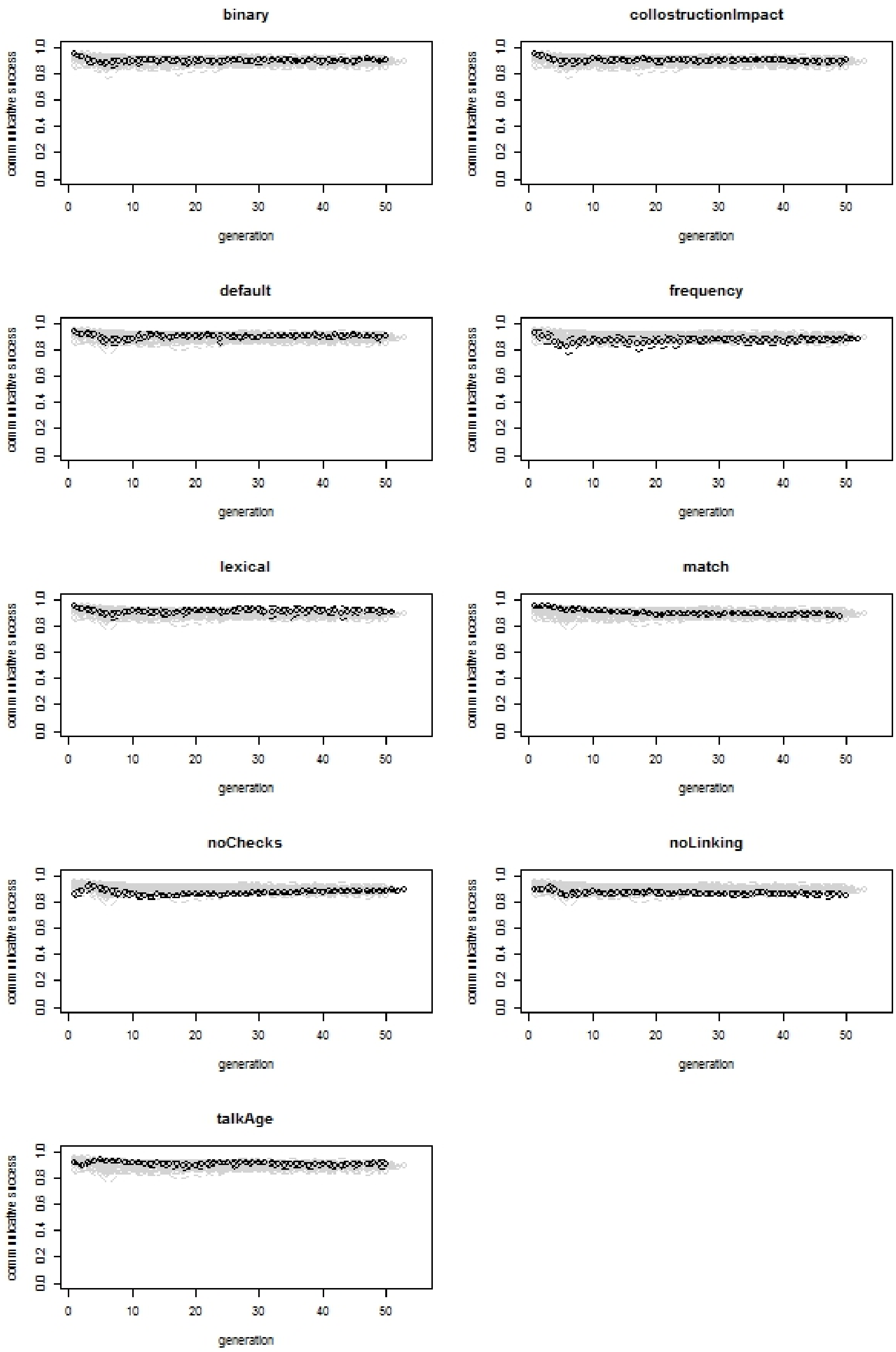

Figure 3: **Communicative success of world type in comparison to other worlds.** Each world type is tested with seven lineages. Dashed lines show standard deviations.