A case of cultural evolution: The emergence of morphological case

Sander Lestrade

Centre for Language Studies, Radboud University Nijmegen

Abstract

Morphological case is a cross-linguistically widespread strategy to mark the role of participants of an event. Using a computer simulation, this paper will show how case can emerge in languages that have no rules of grammar as a result of cultural evolution. For this, only very general cognitive and communicative principles are necessary.

Keywords: argument marking, morphological case, cultural evolution, computer simulation

1. Introduction

Together with word order and verb agreement, case marking is one of the three ways in which languages can make clear the role of the participants in a communicated event.[i] Whereas it left only pronominal remnants in English, in many languages it is actively used as an argument-marking strategy. An example is given from Hungarian in (1), where the accusative marker -*(e)t* marks the book as the thing that is read:

(1) *Anna egy kőnyv-et olvas*.

Anna a book-ACC read

'Anna is reading a book.'

Using a computer simulation, this paper will show how case marking can be modeled as the result of *cultural evolution* (Smith and Kirby 2008; Christiansen and Chater 2008), a variant of evolution that targets the language itself rather than its speakers and works at a relatively fast pace. The main point will be that case does not have to be stipulated as a language prerequisite (the model does not need it in any way), nor as a necessity for communicative success (languages with lexical marking are as successful as "grammatical" languages). Instead, case emerges from very general communicative and cognitive principles.

Van Trijp (2012), one of the very few that studies the development of case via computer simulations, argues that morphological case offers a selective advantage in the communication of events (both for the hearer and speaker) and shows that language users do not need specific innate (syntactic) knowledge for its development. In essence, this paper could be seen as a corroboration of his findings, using a different software environment. Especially his second point is wholeheartedly subscribed to as it is foundational to my research project. There are some important differences, however. Here, it is not assumed that the use of case should be contrasted with the absence of argument marking whatsoever. Instead, case marking develops from the automatization of ad hoc lexical choices to make the argument structure explicit. Thus, case markers are not invented on the spot, as they are in van Trijp's work, but they are recruited from the available lexical means instead and develop into grammatical markers over time only (cf. Heine and Kuteva 2007; Zeevat 2007). Secondly, there is no need for the *shaping* and *multi-level alignment* principles that van Trijp (2012) uses for the development of case markers. According to the first, after successful usage of a marker for some role, the link of that marker with other roles for which it is the dominant expression is also strengthened; according to the second, after successful usage of some marker, other constructions in which the same case is used are rewarded too. Whereas there may be valid reasons for including them in simulations (cf. van Trijp & Steels 2007 for multi-level alignment), they are not necessary in the present set-up.

The work in this paper could be considered complementary to the work of Jäger (2007). Jäger too simulates the development of case systems but rather than studying how certain systems emerge (i.e., taking a historical perspective), he is interested in the stability and spread of the product once available (taking a typological view). Jäger shows that certain case systems are evolutionary stable as they are functionally optimal (maximally economic and sufficiently disambiguating between participants). The question we are interested in here is how such systems can develop in the first place.

2. Modeling event communication and grammaticalization

2.1 The general idea

The R package WDWTW, for *who does what to whom*, is a cognitively motivated, multi-agent model of language usage and change, developed by the author.[ii] The focus, in the present stage, is on event communication, but its open-source and modular architecture in principle allow for many more applications. Also, the design explicitly allows for the check and manipulation of all model assumptions by the user as virtually all assumptions are parameterized.

The agents of WDWTW live in a very abstract virtual world to which none of our real-world concepts apply. They only share with us the presence of a lexicon with referential terms for actions and objects (*verbs* and *nouns*, respectively, anticipating the development of grammar), and very basic cognitive and communicative principles (such as a desire for communicative success, shared attention, recognition of communicative intention, and a principle of least effort; cf. Tomasello 2003). In the absence of a grammar, agents use *proto-principles* to communicate about their world, for example assuming that things that stand together belong together (cf. Section 2.3; cf. Givon 1995 and Jackendoff 2002 for more general discussion).

Two agents find themselves in a situation with a number of randomly generated events going on at the same time. Events consist of actions in which two (non-agent) participants are involved. One of the agents, the *speaker*, has to find a wording for one of the events, the *target event*, that suffices for the other agent, the *hearer,* to single it out.[iii] Words are activated in the mental lexicon of the agent on the basis of both referential match and frequency of usage. The most frequent and best matching words are considered first. The selected words need not describe their referents perfectly, the agent only has to deem them sufficiently distinctive given the situational context. If communication is successful (i.e., if the hearer selects the target event), the usage frequencies of the words produced and heard (which need not be the same) are raised by one, which increases the likelihood of being selected again. If communication fails, the numbers stay the same. After a communicative turn, a new situation is created. If the conversation continues, the speaker and hearer roles change, otherwise, two new speech participants are selected from the population.

The population starts with two agents with a shared lexicon of 2 x 999 randomly generated nouns and verbs. Agents die at 2500 utterances and procreate at the age of 2000, at which point their mixed lexicon is inherited by their two children, resetting all frequency numbers to zero and randomly modifying the meanings of those words that have not been used by the parents until then (the idea being that if words are not used frequently enough, they cannot be learned properly). For present purposes, these simplifications about the development and maintenance of a conventional lexicon seem warranted (for the feasibility of modeling this part of language evolution, cf. Hurford 1989; Hutchins and Hazlehurst 1995; Steels 1997; Kirby 2000).

Initially, all words are fully specified on a number of (abstract) meaning dimensions and have a word length of 7 characters. Over time, however, words can *grammaticalize*, which is modeled as a combination of *erosion* and *desemanticization* (Heine and Kuteva 2007). Words can be ‘‘pronounced’’ sloppily if they are frequent or predictable (cf. Jurafsky et al. 2001 and Balota and Chumbley 1985). In the model, sloppy pronunciation (instantiated as going back in the alphabet for the last letter of a form and deleting it altogether if this is no longer possible) does not lead to a change of lexical representation for the agent using the form. But if the (younger) hearer is still unsure about the form of a word because it has not used it sufficiently frequently yet (viz. 5 times), it will adapt its representation on the basis of what it hears, as a result of which word length may change over time. Also, words may desemanticize. First, they can extend their meaning range incidentally (if the context does not require a more specific description or in the absence of a better expression). Eventually, such an extension may become a standard part of a word’s meaning, as a result of which it becomes more general. In the model, desemanticization involves the progressive removal of the meaning dimensions of a word (cf. e.g. Bybee 2010 for the possibility of such within-generation change). Deletion takes place only after certain *collostruction*-frequency thresholds have been reached (a co-occurrence of two words in a specific constellation, e.g. word *x* as the subject of verb *v*; Stefanowitsch and Gries 2003). For a first dimension to be removed, a word has to be combined with 1% of the relevant predicates. This proportion grows exponentially to 50% for the last dimension to be removed. The dimension at which most variation is attested among the meanings a word had to express is selected for removal.

Before explaining the details of the model further, it is convenient to preview some results first. Consider the expressions for the same event by agents of different generations of the same lineage (the verb is glossed by V):[iv]

(2) a.	*juhibuh*	*mafabep*	*ficesij*	*bekecob*
	JUHIBUH(V)	MAFABEB	FICESIJ JUHIBUHER	
b.	*gipebag*	*dub*	*hapovob*	*ba*
	GIPEBAG	DUB(V)	HAPOVOB	ACTOR

The *a* example is from an agent of the first generation, the *b* example from a descendant of the 100th generation. It can be easily observed that different words are used to express the very same meaning. As the meanings of infrequently used words change at procreation, different words became available to the later agent, which is why the wording in (2) differs between generations, in spite of referring to the very same event. More importantly, the second utterance is much shorter than the first. In this utterance, the word *ba* is in essence a case marker, as will be shown below: a maximally short form with a maximally general meaning that marks its host for its function or type of dependence (Lestrade 2010). It is glossed as *actor* (after Van Valin 1999), as it does not mean much more than 'agentive participant' at this stage. Crucially, this marker was not available yet to the first agent but emerged over time. As a result, the first agent still had to use a lexical expression, which expresses the actor role of the predicate much more specifically (hence the gloss *juhibuher*).

2.2 Meaning representation and comparison

In previewing the results, used was made of such notions as *actor* and *referential match.* Now let us see what these mean in the model. The mental lexicon of the agents is modeled as a list of (randomly generated) forms with values on nine abstract meaning dimensions (cf. the *concept* notion of Gärdenfors 2000). By default the first five dimensions have a low distinctionality (making two or three distinctions only), the other four are much more fine-grained (making 99 distinctions). The number of dimensions and increasing distinctionality are very loosely motivated by linguistic and ontological work: Grammatical systems seem to make use of a few binary distinctions only, whereas the lexicon of course needs to carve up the world into much more detail. Similarly, nodes higher-up in an ontology have to organize the world in major types (e.g., events vs things, concrete vs abstract, humans vs non-humans), whereas nodes further down use a more fine-grained classification (e.g. types of dog). Again, as with virtually all settings of the model, both the number of dimensions and their distinctionality can be manipulated.

One could think of the meaning dimensions in concrete terms (as in natural grammar, e.g. ±animacy, ±concreteness, color, etc.), but the agents only know about abstract numbers. Importantly, the vector representations do not have to form interpretable meanings for us, it is only to provide the agents with a set of meanings to talk about. Just like CAT for us is a concrete, living entity, an animal that is domesticated, from the family of cats, etc.; for an agent *mafabep* is a 0 0.5 0 1 1 0.68 0.22 0.58 0.55, i.e. a 0 on the first dimension 1, a 0.5 on the second, a 0 on the third, etc.[v] What matters for us only is the rules of grammar that develop to organize the relationships between words (and these can perfectly be studied using such abstract semantics).

The action meaning of verbs is similarly specified. In addition, however, verbs have two predicate roles, which are specified on nine dimensions again. Here too, one could translate each role dimension into one that is relevant in the grammar of natural languages (e.g. ±punctual for the action, ±intentional for the actor, and ±affected for the undergoer), but these notions have no meaning in the model.

The match between the referents in the virtual-world and the lexical semantics of the words available in the lexicon of the agent is evaluated by means of vector comparison. The details of this procedure are beyond the scope of this paper; the intuition is that the more similar the values on corresponding dimensions are, the better the match.

2.3 Proto-principles

We can now almost understand why the first agent uttered (2a): It wanted to refer to one of the events in a situation, for which it selected the words that seemed appropriate. Note however that it used four words (just like the later agent), whereas events consist of three ingredients: an action and two event participants. So where did the fourth word come from?

As is clear from the glosses, the non-referential words function as role markers. The use of such markers follows from *proto-principles* of communication*,* very general (non-linguistic) principles whose presence in communication systems preceding language must be assumed in order for the system to be successful and grammar to develop. The principles assumed here are *Check* and *Grouping.*

(Human) communication is bidirectional: Speakers have to take into account their hearers if they want to be understood. In many (theoretical and computional) models, this involves speakers pretending to be hearers to check if they themselves would get the right meaning. If not, they probably have to adapt their utterances (cf. Levelt 1983, Hurford 1989, Steels 2003, Blutner et al. 2006). Economy and predictability are generally considered two important factors in this process. Speakers try to use as economic utterances as possible (i.e. costing little pronunciation effort). Only if this leads to the wrong result, the utterance should be made more explicit (cf. Grice 1975). In the model, the meaning dimensions of nouns and predicate roles correspond to one another. That is, the first predicate-role dimension concerns the same property as the first noun dimension. By comparing these vectors, the degree to which an argument qualifies for its role can be determined (its *typing score*, after Aristar 1997). Whenever the meaning representation of a participant matches the one of the role it performs significantly better than the other participant does, an agent can simply combine the words that refer to the referents. The argument structure then follows from their semantics (i.e., even in the absence of functional word order, we know who's doing what in *book man read*). But if the intended role filler does not qualify better than the other argument and hence the argument structure does not follow automatically (cf. *man boy see*), something extra needs to be done. One solution is to make the role of the participant explicit by naming it using another word that describes this role best. This is what has happened in (2a). Apparently, the agent formulating (2a) thought the argument structure was not sufficiently predictable from the lexical semantics of the event participants alone. To make it clear, it marked the second argument in the utterance, the one at which the ambiguity becomes apparent. As this second argument is performing the actor role of *juhibuh,* the agent used *bekecob* to make it explicit, which, given the two roles of the verb, comes close enough to meaning 'juhibuher'.[vi] This whole procedure of checking intended role distribution and adding markers when necessary is operationalized as the proto-principle *Check.*

Now adding words to make roles explicit wouldn't help much if it wasn't for the second proto-principle *Grouping* (cf. Givón's 1995:188 *proximity-relevance* constraint). The idea is that things that stand together probably belong together. The hearer has to understand that *bekecob* is not meant as a referential expression itself, but as an (ad hoc) role marker that needs to be combined with another word. The task for the grouping algorithm is to find the grouping analysis (after identifying the verb) that makes most sense in the communicative context. For our example, in which *juhibuh* is recognized as the verb, this means that there are five possibilities: either *mafabep, ficesij,* or *bekecob* is left unanalyzed, the other two being the arguments (but leaving elements unanalyzed is penalized), next *ficesij* could be the marker of *mafabep*, or *bekecob* that of *ficesij*. Using the speech situation in which referents can be found for all words except *bekecob*, the hearer decides to analyze this word as a role marker. As its meaning comes closer to the prototypical 'juhibuher' than to 'juhibuhee' (which is why the speaker choose it in the first place), the hearer interprets *ficesij* as having the actor role.

4. The emergence of case

As we saw above, both agents agree that something extra needs to be done as the participants fall short of their respective roles, because of which the distribution of roles does not follow automatically. Whereas he first generation uses *bekecob* to mark the actor participant for its role, the second uses *ba*. Note that it is only accidental that in both cases the actor argument is marked: Ambiguity is resolved on the second argument only, and in both utterances the actor argument happens to come last (word order being free). Had the undergoer come second, the first agent would have used *dabibof* to mark it as such, whereas the later agent would have used *bab* or *be*.

The formal difference between the two solutions is clear: Whereas *bekecob* is of full length (all words initially are 7 characters long), *ba* is maximally short (the minimum length is 2 characters and *a* is the "cheapest" vowel; cf. the erosion procedure described in Section2.1).[vii] Only 23 (plus 11) out of 999 noun (plus 999 verb) forms became this short after 100 generations. Moreover, *ba* is among the most desemanticized words, having lost three of its meaning dimensions, whereas the vast majority of words, viz. 860, is still fully specified.

By the 100th generation, then, *ba* became the default way to mark the actor. Unlike referential expressions, which have to single out their referent from many distractors, the only thing a role marker has to do is distinguish between the two roles of a verb. As a result of this, they need not be very precise (i.e., even if a word is not much of a 'juhibuher', if it's even less of a 'juhibuhee' it can be used to mark the actor role of *juhibuh*). And as a result of this, markers that are (frequently) used before, are more likely to be found good enough to be used again (activation in the mental lexicon being partly dependent on usage frequency). Next, since meaning dimensions along which a lot of variation is attested are removed from the meaning representation of words, a meaning core remains that captures whatever most actor roles share. (The two other markers mentioned above, *bab* and *be*, are still in competition for marking the undergoer.) Now if we think of case markers as maximally short forms with maximally general meanings that mark their host for its function or type of dependence (Lestrade 2010), *ba* in essence is the (ergative) model equivalent of the case marker in (1).

5. Conclusion and discussion

Protolanguages turn out to be perfectly capable of communicating events by means of ad hoc solutions (both (2a) and (2b) are correctly interpreted by the hearer) and there is no improvement nor decay of communicative success due to the development of case markers. Thus, there is no real communicative reason for case markers to develop. Instead, they are the by-product of cultural evolution as a result of a complex (but understandable) interplay of factors: Early speakers check if they will be understood and add lexical role markers if communicative failure is imminent. Roles need to be marked frequently and can be kept apart relatively easily (without being very precise). Since frequency plays a role in word activation, good-enough solutions are more likely to be used again, further increasing their frequency. Frequently used words are pronounced sloppily and hence erode as they may be learned "wrongly". In addition, promiscuously combining words loose meaning specificity, a process which also propels itself once started (as vaguer words combine even more easily). Thus, over time, case markers (short forms with general meanings marking dependency) may develop from popular lexical ad hoc solutions. Interestingly, whereas case marking arguably leads to a more efficient communication system eventually, this thus need not be its selective advantage.

There are many issues that could not be discussed for reasons of space. Most importantly, the parameter space of the model should be more systematically explored. For example, the thresholds for grammaticalization and the availability of alternative (emergent) disambiguation principles such as word order and head marking influence the development of case marking. Also, a number of simplifying assumption were made (in addition to the ones already mentioned in passing): there is no coexistence between lexical and grammaticalized meaning (within agents) and social dynamics (who's talking with whom) are not properly modeled yet. Finally, the link between the largely semantic markers discussed here (marking actor and undergoer) and the structural cases marking subject and object, as in (1), needs to be developed. Although the idea is simple (most undergoers are internal arguments, because of which the undergoer marker may be reanalyzed as internal marker), its model implementation is still necessary.

Acknowledgments

I would like to thank Helen de Hoop, Ruti Vardi, and two anonymous reviewers for valuable comments that helped to improve the paper. Also, thanks to the Netherlands Association for Scientific Research (NWO) for their financial support (grant 275-78-001, *The exaptation of argument marking*).

References

Aristar, A. R. 1997. "Marking and hierarchy. Types and the grammaticalization of case markers". *Studies in Language* 21:2.313–368.

Balota, D. A. & J. I. 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.

Blutner, Reinhard, Helen de Hoop & Petra Hendriks. 2006. *Optimal Communication*. Stanford: CSLI.

Bybee, Joan. 2010. *Language, Usage and Cognition*. New York: Cambridge University Press.

Christiansen, M. H. & N. Chater. 2008. "Language as shaped by the brain". *Behavioral and Brain Sciences* 31:5.489-509

Gärdenfors, Peter. 2000. *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: MIT.

Givón, Talmy. 1995. *Functionalism and grammar*. Amsterdam and Philadelphia: PUB.

Grice, H. Paul. 1975. "Logic and conversation". *Syntax and Semantics*, Vol. 3, *Speech Acts* ed. by P. Cole and J. L. Morgan, 41–58. New York: Academic Press,

Heine, Bernd & Tania Kuteva. 2007. *The genesis of grammar. A reconstruction*. Oxford: Oxford University Press.

Hurford, J. R. 1989. "Biological evolution of the saussurean sign as a component of the language acquisition device". *Lingua* 77:2.187-222.

Hutchins, Edwin and Brian Hazlehurst. 1995. "How to invent a lexicon: The development of shared symbols in interaction". *Artificial societies: The computer simulation of social life* ed. by N. Gilbert and R. Conte, 157–189. London. UCL Press.

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

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

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001. "Probabilistic relations between words: Evidence from reduction in lexical production". *Frequency and* *the emergence of linguistic structure* ed. by J. Bybeel and P. Hopper, 229–255. Amsterdam/Philadelphia. John Benjamins.

Kirby, Simon. 2000. Syntax without natural selection". *The evolutionary emergence of language* ed. by C. Knight, M. Studdert-Kennedy & J. R. Hurford, 303–323. Cambridge, UK.

Lestrade, Sander. 2010. *The space of case.* PhD thesis, Radboud University Nijmegen.

Levelt, W.J.M. 1983. "Monitoring and self-repair in speech". *Cognition* 14.41-104.

Nichols, J. 1986. "Head-Marking and Dependent-Marking Grammar". *Language*, 62:1.56-119.

R Core Team 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Smith, K. and S. Kirby. 2008. "Cultural evolution: implications for understanding the human language faculty and its evolution". *Phil. Trans. R. Soc. B* 363.3591–3603.

Steels, Luc. 1997. "Constructing and sharing perceptual distinctions". *Machine Learning: ECML-97*, 4–13. Springer.

Steels, Luc. 1999. *The talking heads experiment*, volume 1. Brussels: VUB Artificial Intelligence Laboratory.

Steels, L. 2003. "Language re-entrance and the inner voice". *Journal of Consciousness Studies* 10:4-5.173–185.

Stefanowitsch, A., & S. T. Gries. 2003. "Collostructions: Investigating the interaction of words and constructions". *International Journal of Corpus Linguistics* 8:2.209-243.

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

van Trijp, Remy. 2012. "The evolution of case systems for marking event structure". *Experiments in cultural language evolution* ed. by L. Steels, 169–206. Amsterdam and Philadelphia. John Benjamins.

Van Valin, Robert D. 1999. “Generalized semantic roles and the syntax-semantics interface”. *Empirical issues in formal syntax and semantics* 2 ed. by F. Corblin, C. Dobrovie-Sorin & J.-M. Marandin, 373–389. The Hague: Thesus.

Zeevat, Henk. 2007. “Simulating recruitment in evolution”. *Cognitive Foundations of Interpretation* ed. by G. Bouma, I. Kraemer & J. Zwarts, 175–194. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

[i] The notion of case is used in a very liberal sense here, as the best-known instantiation of *dependent marking* (Nichols 1986). The phonological distinction between various forms of dependent marking (adpositions, clitics, and affixes) is irrelevant for present purposes.

[ii] As the package is still being developed, it is not yet available in the CRAN archive. Meanwhile, the codes are available from the author upon request.

[iii] This set up, taken from Luc Steels' robotic simulations (cf. Steels', 1999 ''guessing game'') but then without repair in case of failure, was suggested to me by Simon Kirby (p.c.).

[iv] In normal runs, it is extremely unlikely that agents from different generations have to express exactly the same event. For this, deceased agents have to be resurrected, something the model allows for.

[v] For computational reasons, values are rescaled to the 0-1 range.

[vi] Note that bidirectional reasoning was already applied in the process of selecting the referential words described above: if a word is not sufficiently distinctive given the context, it will not be used.

[vii] As agents will not pronounce forms sloppily if lexical ambiguity arose from it, both *bab* and *be* forms are blocked from further erosion because of the presence of *ba.*