How to explain the origins of complexity in language

Steels, Luc; Beuls, Katrien

Published in:
Complexity in Language

DOI:
10.1017/9781107294264.002

Publication date:
2017

Document Version
Peer reviewed version

Explaining the Origins of Complexity in Language

A Case Study for Agreement Systems

Luc Steels[1,2] and Katrien Beuls[3]

[1] ICREA/Institut de Biologia Evolutiva (UPF-CSIC), Barcelona
[2] Sony Computer Science Laboratory, Paris
[3] AI Laboratory, Vrije Universiteit Brussel

Abstract

The different forms of complexity observed in human languages (inventory complexity, form complexity, processing complexity, learning complexity and population-level complexity) can be explained by taking a functional and evolutionary point of view. The functional viewpoint insists that language is a tool for communication and hence language users expand their language systems to increase expressive power while keeping in check cognitive effort, for example, by re-using existing elements for new purposes or by limiting combinatorial search. The evolutionary viewpoint emphasizes that a language community has to collectively find within the open-ended space of possible languages a candidate that satisfies their purposes. This can be explained in terms of (cultural) selection. Language users employ various strategies to construct and maintain their language and those strategies that contribute to expressive power and the damping of cognitive effort should be maintained and variant strategies discarded. This paper reports on a number of computational simulations of multi-agent models in which these functional and evolutionary viewpoints have been explored. Strategies for the emergence and evolution of grammatical agreement systems are used as case study.

Keywords: Language complexity, language dynamics, agent-based models, grammar emergence, language evolution.

1. Introduction

It is useful to make a distinction between five different types of complexity in language:

1. The *complexity of the inventory* is about the number of conceptual, phonological, lexical and grammatical building blocks in use by an individual or by a particular language community: The number of phonemes, the number of concepts (for example color categories, action categories, spatial relations, temporal relations, etc.), the number of lexical items (words and morphological elements), the number of grammatical categories (syntactic classes, cases, classifiers, types of phrase structures), and the number of grammatical constructions.

2. *Form complexity* is about the size of the linguistic forms that make up an utterance. It can be measured in terms of statistics over the length of words and the length of utterances.

3. *Processing complexity* is defined as the cognitive effort involved in parsing and producing utterances: How much memory is needed? How many processing steps? How much combinatorial search is unavoidable? How much ambiguity is left before semantic interpretation? Processing complexity depends on the architecture of the language processing system, the complexity of the inventory, the complexity of forms, and the ecological complexity of the environment in which language users operate.

4. *Learning complexity* concerns the amount of ambiguity that learners face when acquiring new words or constructions. Is it always possible to uniquely guess unknown meanings and functions, or does a search space get generated? How big is this search space?

5. *Population-level complexity* is concerned with the properties of the evolving communal inventory as a whole: How much variation is there in the population with respect to the usage or knowledge of linguistic forms? Or conversely, what is the coherence, i.e. the degree of sharing? What is the resilience of particular constructions in the process of cultural transmission? How intense is language change?

The central thesis of the present paper is that human languages evolve in such a way as to minimize complexity at all these levels while providing enough expressive power to handle all the meanings relevant to the community. Minimizing complexity is necessary to keep the language viable, otherwise it would become too complex for regular usage and would no longer be learnable (Hawkins, 2005). Minimizing complexity does not happen by intelligent design. There is no central committee that can oversee and regulate the use of English, Chinese or any other language. Speakers have only limited knowledge of the language used in their population and have no way to directly influence others. The alternative to intelligent design is a selectionist process (Mufwene, 2001; Steels, 2012b): Language users unavoidably generate variation and complexity at all levels, often because they do not know which alternatives already exist in the population. At the same time, each individual adopts strategies that have the aggregate effect of minimizing complexity of the language as a whole. Variants for concepts, words and constructions compete and those variants that lead to higher communicative success and less cognitive effort have a higher chance of survival. There is no optimal solution, and so a language keeps moving around in a linguistic fitness landscape, sometimes optimizing one aspect (e.g. minimizing form complexity) while relaxing another aspect (processing complexity). How do language users minimize complexity? There are many mechanisms and they depend on the type of complexity we are looking at:

1. *Damping inventory complexity.* Language is an open system. New meanings constantly come up and need to become expressible. Routine expressions tend to loose their force and are then replaced by new ones. Consequently, there is a steady renewal in the inventory of a language: New words are invented or existing words used for new meanings. New grammatical constructions are introduced, sometimes based on new paradigms or on the application of a paradigm to new cases. Existing constructions may also be coerced into new uses. At the same time, inventories must be kept within bounds because larger inventories take up more memory, increase access time, and require longer time for learning.

How are the inventories kept in check? Many words or constructions simply get out of fashion after a while. They disappear from usage and are forgotten, particularly after a few generations. At the same time, there is a general tendency to reuse as much as possible existing forms even if they have already existing functions. If a new invention is based on the exaptation of an existing word or construction in a slightly different context, then there is a higher chance that the hearer might guess this new meaning than if a radically new invention is made. Hence the exapted invention has a higher chance to propagate and survive in the communal language and it contributes to keep the language inventory in check.

2. *Damping form complexity.* Novel words tend to be long, but, once introduced, we see that words begin to erode phonetically due to articulatory optimization or errors, reducing both the length of words and the length of utterances in which they appear. Some words end up being morphemes, then clitics and later affixes. A similar form optimization process happens at the level of constructions. The first time new meanings are expressed is usually done in an elaborate, circumscriptive way. But with routine usage, the combination of constructions that was used to build a more elaborate phrase gets collapsed into a single construction to achieve faster processing, a process usually called chunking. Words within such a chunk may then start to disappear and the phrase may progressively become idiomatic.

3. *Damping processing complexity.* Strategies for reducing inventory size and form complexity may lead to syncretism (the same form gets multiple meanings) and hence syntactic ambiguity. Syntactic ambiguity leads in turn to combinatorial search, which needs to be kept in check because it causes an exponential increase in memory and processing time and hence fluent rapid speaking and listening becomes more and more difficult. Inventory reduction by reuse of existing forms may also lead to semantic

ambiguity because the forms in the utterance no longer unequivocally communicate the intended meaning. Processing complexity can be damped by adding additional grammatical structure. For example, as discussed later in this paper, grammatical agreement is a way in which language users signal which words belong together in the same sentence and hence it reduces the combinatorial complexity of parsing. More generally, we argue that the reduction of cognitive effort in parsing and interpretation is one of the main motivating forces for the origins of grammatical structure.

4. *Damping learning complexity.* Language learners need to guess the meaning and function of unknown words or constructions. The more possibilities there are, the more hypotheses the learner has to consider. The learning task gets more manageable if the learner can make strong use of context, which reduces the set of possible meanings, and if the speaker, acting as tutor, helps by scaffolding the complexity of utterances and by correcting wrong uses of words and constructions. At the same time, the structure of language itself can help. For example, syntactic structure may not only decrease the complexity of parsing but also help to constrain the possible meanings of an unknown word, leading to syntactic bootstrapping. For example, the adjectives in a nominal phrase are typically ordered based on semantic classes and hence this ordering can help constrain the meaning of an unknown adjective.

5. *Damping population-level complexity.* Individuals in a population will never share the same inventory because language learners have to independently acquire the language based on their own history of interactions with others and their own needs. Hence we see tremendous variation within language communities. On the other hand, language variation is detrimental to the efficiency of a language because language users need to be able to parse and interpret expressions that deviate from their own usage and possibly even store these variations. Instead of learning a single language they have to learn a multitude of overlapping languages. So we need a strong force that dampens population level complexity and ensures that idiolects become shared into a common language. One of the important discoveries of the past decade is that a selectionist dynamics can be realized if each language user associates with each element in his or her inventory a score, which reflects the best choice while producing or interpreting an utterance. This score is updated based on the use and success of the variant. One possible strategy (known as lateral inhibition) works like this: When a variant is part of a successful communication, its score is increased and the score of its competitors decreased. And when it is part of a failing utterance, the score is decreased. Many computational simulations (one of them shown later) have demonstrated that this mechanism effectively allows a population to align their inventories.

2. Agent-based modeling

There are several ways to study language dynamics: linguistic analysis that deconstructs the kinds of rules that must intervene in the parsing and production of utterances, studies in historical linguistics which show how a language has been shaped and reshaped over time by its users, research into language typology and sociolinguistics, which maps the diversity and variation in language, psychological experiments that measure delays in response to utterances or difficulties of learning, and neuroimaging experiments which measure brain activation while parsing, producing or learning utterances. All of these approaches are valuable. In our team, we have been using a complementary method, namely agent-based modeling (Loreto et al., 1997; Smith et al., 2003; Steels, 2011). It can be used to pin down precisely by what kind of cognitive mechanisms, strategies and cultural processes linguistic complexity arises and how it gets damped.

Agent-based models assume a population of linguistic agents that engage in situated embodied interactions, called *language games*. A concrete agent-based model defines a set of possible situations and an interaction pattern between agents that contains both verbal and non-verbal aspects (such as pointing). It contains furthermore a set of mechanisms that the agents in a population are given initially, prior to any interaction and strategies using these mechanisms to build up a new language system from scratch. Then a series of games is being played between randomly chosen agents from the population. Each agent can play both the roles of speaker and listener. As a side effect of a game, the participating agents may expand their conceptual or linguistic inventory, or adapt it, based on the outcome of the game. We can track different measures, both for individual agents and for populations of agents, in order to tell us whether a model works, in the sense of whether the desired linguistic structures indeed arise in model simulations. If they do not, the mechanisms initially put into the agents are insufficient and they are changed for the next experiment until we know what strategies generate the phenomena we want to explain.

For example, in experiments trying to explain how a set of color terms and color concepts could emerge, agents would initially be endowed with concept formation mechanisms and strategies for inventing, acquiring or aligning associations between words and color concepts (Steels and Belpaeme, 2005). At the start of the experiment there is neither a shared color language nor any color concept inventory, but if we see after a number of games that a shared color language emerges, we have evidence that the given strategies (plus the interaction patterns and ecological conditions) explain the emergence of color terms and the color concepts they express.

An agent-based model typically focuses on one aspect of language. Examples can be found in a recent volume that contains case studies for color terms, names for actions, a tense-aspect system, quantifiers, expression of information structure, phrase structure, case grammar, grammatical function, and others (Steels, 2012a). Each time a particular language game is chosen that brings out the aspect of language being studied. For instance, if we want to study the emergence and evolution of color terms we obviously need an environment with objects of different colors and the color should matter in the interaction between the agents.

Agent-based models have numerous advantages compared to verbal theorizing and it is therefore surprising that this methodology is still controversial in language evolution research (even though agent-based models are widely accepted in sociology, biology, economics and many other scientific disciplines):

- Agent-based models require a comprehensive mechanistic model of language processing, which is usually lacking in linguistic or psychological research. Most often grammars are not formally represented, and if they are, the formalization is often not in a form that can be used by processing algorithms. Various learning strategies are assumed but they are only described vaguely and not operationalized, so that we cannot know whether they are effective. Agent-based models can only be built with effective computational models of language processing and because these models need to cope with emerging and evolving languages, they challenge several dogmas in formal and computational linguistics (Steels, 2012b). In that sense, agent-based models push not only the state of the art in language evolution but also in linguistic theory in general.

- With agent-based models it is possible to measure precisely the complexity of a given model at each of the different levels of complexity described earlier. We can do repeatable experiments and examine the statistical distribution of the results. It is possible to selectively add or take out mechanisms or change a strategy so that we can determine the causal relationships between mechanisms, strategies or parameters of the model and the structure of an emergent language system.

Agent-based models are occasionally criticized because the strategies and processing mechanisms are stated at an abstract, mathematical/computational level, but this is actually the case for every scientific model. They are also criticized because such models are not validated by psychological observations or neuro-imaging data. The latter is true. But first of all we are still far removed from effective models of language processing that have been or even could be validated this way - given the precision of current neuro-imaging experiments. And second, cognitive psychologists have been replicating similar conditions as used in agent-based models with remarkable results (Galantucci, 2005; Selten and Warglien, 2007). Agent-based models are a huge opportunity for psychologists. They provide possible models of language processing and language evolution, which have been proven to work, and hence it makes sense to validate them against psychological evidence. There is a third point. The observations made by historical linguists are in fact the ultimate empirical target of agent-based models of cultural language evolution. Historical linguists have identified many of the processes that are active in language emergence and change, such as shifting in the syntactic categories of words, phonological erosion of recruited word forms, damping of synonymy, and so on (Heine and Kuteva, 2007). Similar processes have been identified in the formation of creoles (Mufwene, 2008). The main goal of agent-based models is to show the mechanisms and strategies behind these empirically attested phenomena. Of course we cannot directly replicate precisely the historical evolution of Old English into Middle English, but we should be able to demonstrate by what strategies and cognitive mechanisms the underlying grammaticalization phenomena can happen in principle, for example, how new syntactic categories may arise.

3. Case study for agreement systems

The best way to illustrate both the theory of language evolution by linguistic selection and the methodology of agent-based modeling is to look at concrete examples. We will do this now for grammatical agreement, building further on agent-based selectionist models reported earlier (Beuls and Steels, 2013). Grammatical agreement occurs when two linguistic units (typically a word or a phrase) receive a marker that indicates their relatedness (Lehmann, 1988). For instance, in the utterance *une belle fille* 'a pretty girl' all the words are marked with an *-e*, thus indicating the feminine gender of the referent. Agreement systems can be very complex and messy and many linguists (and second language learners) have been puzzled why they are there. We now look at the different types of complexity discussed earlier and show how specific strategies, adopted by languages with agreement systems, can dampen them.

3.1 The Language Game

In the current experiment, agents play a game of reference (known as the Naming Game), which has been used in many earlier investigations of language dynamics (Dall'Asta et al., 2006; Steels, 1996). All agents can play the role of speaker or hearer and have an equal chance of playing a game. The agent chosen as speaker goes through the following steps: (i) The speaker selects a subset of the objects in the current situation to act as the topic of the utterance. (ii) The speaker looks for a distinctive combination of properties for each of the objects in the topic, where a distinctive combination is a set of properties which are true for the object but not for any other object in the current situation. (iii) The speaker retrieves the minimal set of words in the vocabulary that cover the chosen properties, which implies that words with the largest coverage are preferred. The speaker then utters these words. Given the experimental parameters adopted in the simulations (discussed in more detail in (Beuls and Steels, 2013)), there is ±25% chance that 1, 2, or 3 word utterances are produced but there are also 4-word (15%) and 5-word (5%) utterances. Although there is unavoidably a sequential ordering to the words, this does not carry any meaning, i.e. agents use a word-order free language.

Next, the agent chosen as hearer goes through the following steps: (i) The hearer looks up the words in the vocabulary and thus reconstructs what properties have been communicated by the speaker. (ii) The hearer identifies which objects in the present situation satisfy these properties and points to them. The game is a success if the objects pointed to by the hearer are those initially chosen by the speaker. The game fails if this is not the case or if the utterance remains semantically ambiguous, i.e. if there is more than one possible interpretation that fits with the current situation model.

3.2 Processing complexity

Although this language game looks deceptively simple, there is the potential for a combinatorial explosion and semantic ambiguity. The hearer does not know how many objects the speaker is talking about and the utterance does not communicate which words are about the same object. Hence, all possible combinations must be tried by the hearer to find those that fit with the current situation. This obviously implies a very high degree of processing complexity.

Processing complexity can be measured in different ways and it always depends on the architecture of the language processing system that one has implemented. It could be the nodes in the search tree during production or parsing of a particular sentence, the number of constructions needed to build or understand an utterance, or the 'accessibility' of the constructions that are used to parse or produce the utterance. Frequently used constructions can be primed while others are harder to recall. All of these measures tell us something about the cognitive effort of the speaker or the hearer.

In the case of agreement one relevant complexity measure is the number of hypotheses that remain after parsing a particular utterance. For instance, if there would be neither word order nor any other kind of grammar and I would say "brown blue cat chair big", I could have meant:

- (cat) (big blue brown chair);
- (brown cat) (big blue chair);
- (big cat) (blue brown chair); or
- (blue cat) (big brown chair)
- (big brown cat) (blue chair)
- (big blue cat) (brown chair)
- (big blue brown cat) (chair) ... etc.

And this still assumes that there are only 2 objects. If an utterance can be about more than one object, then the set of possible combinations is still much larger. More generally, the number of possibilities *Bn* is equal to the number of partitions of the set *D* of words in an utterance of size *n*. *Bn* is known as the Bell number and can be defined using the following equation (Bell, 1938):

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$$

with $B_0 = 0$ and $B_1 = 1$. *Bn* grows exponentially with the number of words. It means that the sentence you are now reading (which contains 20 words) generates 51,724,158,235,372 partitions and hence possible interpretations.

However, now suppose that the agents adopt a strategy to use agreement markers, for example, they could simply add an arbitrary morpheme to every word that refers to the same object. An example utterance could now be "brown-ki blue-ba cat-ki chair-ba big-ba". It is now directly clear which words belong together and the combinatorial explosion disappears. Agreement markers are therefore an efficient means to dampen processing

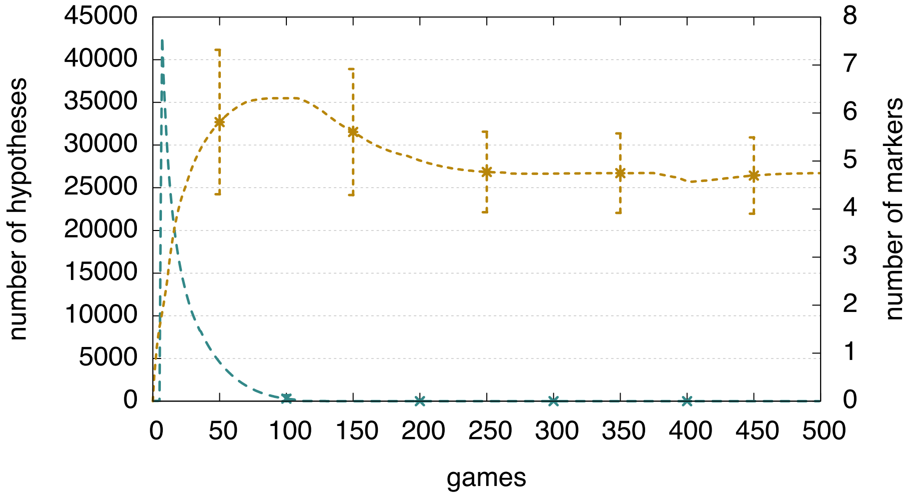


Figure 1: Results of an agent-based simulation, with a population of 10 agents playing 500 language games in total, which means an average of 25 per agent. The x-axis shows the number of games played, the left y-axis the number of hypotheses left after parsing and the right y-axis the number of markers. We restrict the number of objects to 3 so that the optimal number of markers is 3. When agents use the agreement marker strategy, the number of hypotheses decreases rapidly thus damping drastically processing complexity.

complexity as all the combinations that are evoked by the Bell number are hereby reduced to a single one.

This is demonstrated in an agent-based experiment, in which a population of agents starts without an agreement system but with a strategy to invent new markers when needed (as speaker) or adopt them (as hearer). We see in Figure 1 that the number of hypotheses gets drastically reduced as soon as the agents share a sufficient set of markers.

So this example illustrates that the strategy for adding agreement markers dampens processing complexity and hence that it should receive positive selection. If a population of agents uses such an agreement strategy it needs less cognitive effort in parsing and interpreting utterances then if it would not, and this explains why we find agreement systems in human languages.

3.3 Population-level complexity (language variation)

When speakers invent new markers, there is unavoidable variation because agents are not supposed to have a telepathic general overview of all interactions happening throughout the population. Agents end up with a working system, but there are many more markers than is absolutely necessary and hence agents have to learn many more markers. This in turn slows down the emergence of the agreement system and makes it harder for new agents coming into the population to acquire the existing set of markers.

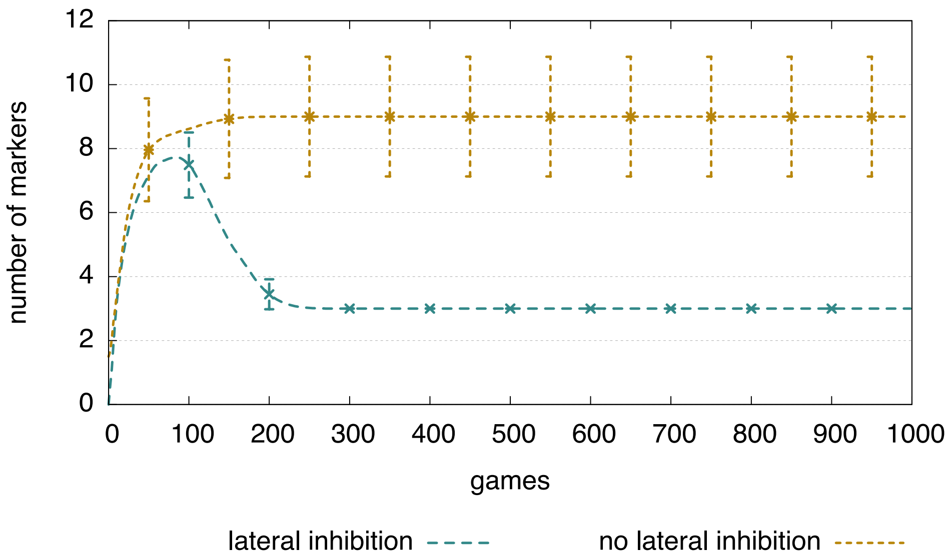


Figure 2: Comparison of a population that uses lateral inhibition and one that does not. The use of lateral inhibition gives a selectionist advantage because it leads to a smaller marker inventory at the population level, i.e. less unneeded variation.

This unnecessary language variation can be damped when agents use the strategy of lateral inhibition as part of the marker strategy discussed in the previous paragraph (de Vylder and Tuyls, 2006; Steels, 1998). Agents maintain a score σ_m between 0.0 and 1.0 for every marker *m*. The initial score of a new marker is $\sigma = 0.5$. The hearer (but not the speaker) increments this score whenever a marker *mi* appears in an utterance and decreases the score of all other non-used markers *m j* using the following equations with alignment rate $\gamma = 0.2$:

$$\sigma_{m_i} \leftarrow \sigma_{m_i}(1-\gamma) + \gamma$$

$$\sigma_{m_j} \leftarrow \sigma_{m_j}(1-\gamma)$$

When choosing which marker to use, the speaker prefers the marker with the highest score that was not yet used in the same utterance. This establishes a positive feedback loop between usage and marker preference, which leads to a shared minimal marker system. The resulting process is similar to self-organizing processes found in natural systems in which large-scale structures arise from local interactions through random behavior influenced by positive feedback loops (Camazine S. et al., 2001).

The effect of using lateral inhibition is illustrated in Figure 2. When agents simply invent markers or acquire them from others, the number of markers becomes stable with an inventory of 10 markers. When they use lateral inhibition as part of the agreement strategy, they end up with the optimal number of 3 markers (because utterances in this experimental run have been limited to contain maximally 3 objects). When there are more than 3 objects, the population would settle on more markers but always the minimal number needed.

This example demonstrates again that the adoption of a particular strategy can lead to a more efficient language system in the sense that it dampens complexity along a particular dimension. In this case, lateral inhibition dampens variation in the population, so that each agent needs to keep fewer markers in memory and new agents need to learn fewer markers.

3.4 Learning complexity

Human languages do not use purely formal markers. For example, the Swahili marker *ki-* used as an agreement marker in the phrase *ki-kapu ki-kubwa ki-moja* (ki.SG-basket ki-large ki-one) (Welmers, 1973) is used for the class of inanimate objects, which contains artifacts such as baskets. The Latin marker *-arum* expresses plural, feminine and genitive. Meaningful markers make it possible to express more meaning with fewer linguistic forms, an important economizing principle of language. If a word already expresses the meaning of a marker, the marker is no longer needed and can be left out. For example, “fille” (Fr. girl) already expresses feminine and therefore does not need an extra marker of gender. Moreover, a marker can introduce additional meanings on top of the meaning supplied by words. This is the case with the Latin *-arum*, which not only carries out the agreement function but also conveys that the referent is plural and genitive. Hence, using meaningful markers is a better strategy from the viewpoint of damping the size of the inventory and the size of an utterance.

Studies in grammaticalization have shown that in the historical record, agreement systems initially arise by reusing existing words as markers (Fuss, 2005; Givón, 1976). Why would that be? We argue that this dampens learning complexity. If an arbitrary label is used as meaningful marker, then the meaning of this marker needs to be guessed and agreed upon, and there is a higher chance that different agents introduce different markers. But if an existing word is used, then the hearer can make a much better guess of the possible meaning of the marker and hence the search space for learning the marker is much smaller. In fact the learner only has to become aware that the word is now used not only lexically but also grammatically.

We have done another experiment in the context of agreement to explore this phenomenon. Learning efficiency is measured as follows. Let the *inventory size* I_g be the total number of distinct marker constructions invented by the whole population up till game g and U_g be the number of markers being in used for the same window, then the *learning efficiency* $S_g = U_g/I_g$ captures how well superfluous inventions could be avoided within this interval. This gives an indication of the learning efficiency because the fewer constructions learners add to their inventory (as a possible hypothesis of the shared language), the lower S_g will be. Figure 3 compares two strategies: One in which new markers are just random symbols (the no-reuse strategy) and a second strategy where new markers are based on existing words (the reuse strategy). We see that the learning efficiency is considerably different: Only 20% of the constructions are still circulating with the reuse strategy whereas 60% of the constructions survive in the case of no-reuse (Figure 3). This is another example where a particular strategy allows agents to dampen complexity, in this case along the learning dimension.

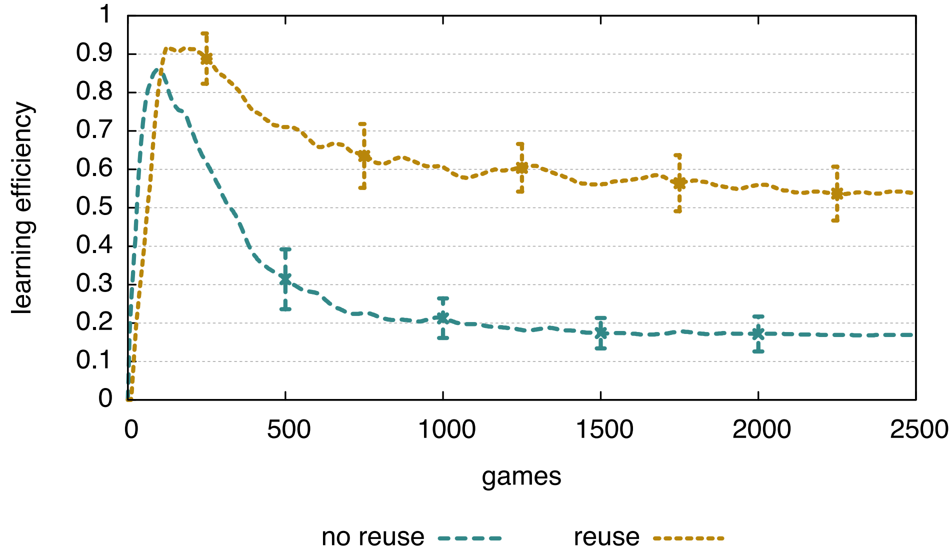
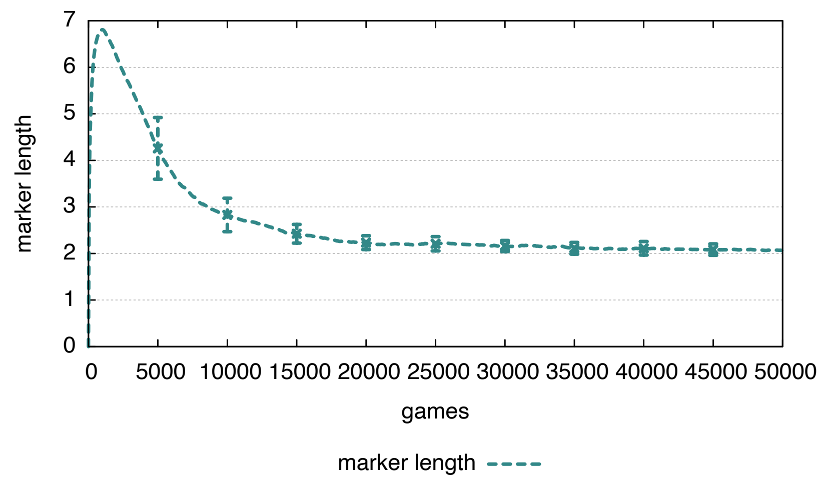


Figure 3: Experiment comparing the learning efficiency S_g between a strategy that reuses existing words as meaningful markers and one that does not. We see that the first strategy is much more efficient in the sense that learners generate fewer hypotheses about the possible meanings of markers that are then later not used.

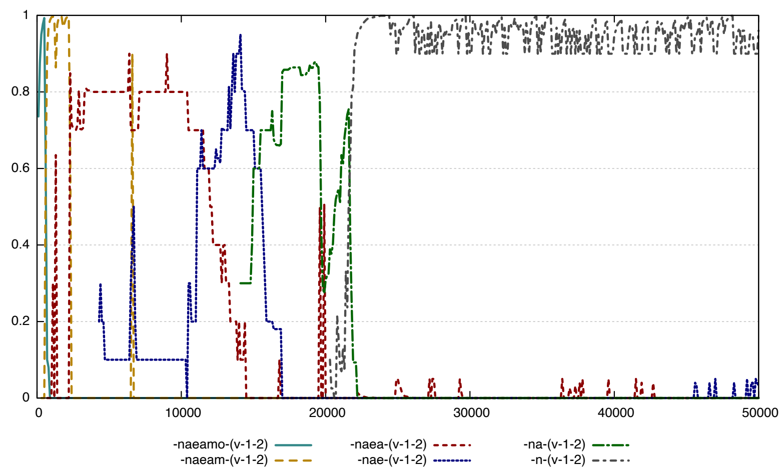
3.5 Form complexity

Another phenomenon that is observed quite often in language evolution is that the form of a word gets shortened and thus its form complexity reduced. This has also been observed with agreement markers, which are often shortened so that the original word is no longer recognizable. We have run therefore additional experiments in which agents use this strategy (in addition to the strategies discussed earlier). Speakers optimize articulation by leaving out the last consonant or vowel of a marker with a certain probability $\varepsilon = 0.1$. Hearers are flexible enough in their parsing of markers to recognize that a truncated form is a variant of an existing marker, as long as it deviates for only one consonant or vowel. This maintains an adequate level of communicative success and does not diminish the effectiveness of markers to cut down combinatorial complexity and semantic ambiguity. But how can we explain that a variant might itself become the norm and in turn become the subject of further optimizations?

Figure 4 shows the outcome of a computer simulation of this phonological reduction strategy. An agreement system based on meaningful markers is emerging using the meaningful marker strategy. But after agents reach a stable level of performance (in the experiment this is typically after 200 games per agent), they occasionally introduce phonological reductions with probability $\varepsilon = 0.1$ and this leads to the erosion of the original markers. Figure 4a shows that the average marker length is decreasing from an average of 7 to 2 consonants and vowels, without affecting performance. There is greater variation Vg in the population because there are always different variants of the same marker in use, but this generally does not have an impact because agents are able to recognize them as a variant of their own norm.



(a)

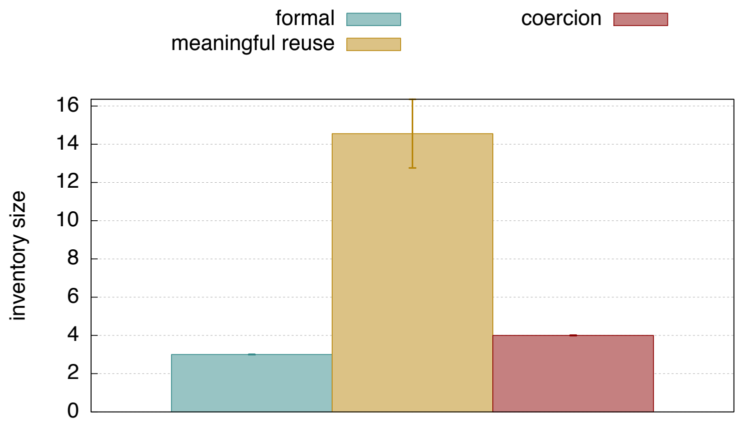


(b)

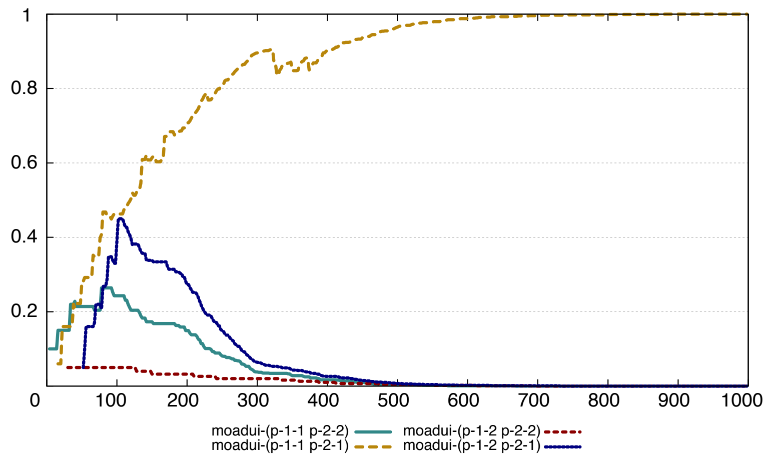
Figure 4: (a) Overall reduction of form complexity in a population of agents over 50 000 games. (b) The "-naeamo" marker (expressing feature v-2-1) starts off with 6 letters and erodes down to a single letter "-n". Yet, after 25 000 games the furthest eroded marker is sometimes alternated with longer versions of the original marker ("-naea" and "-nae"). This scenario occurs when one agent in the population hasn't yet adapted to the conventional marker use of a feature that is not frequent in the world.

3.6 Inventory complexity

Agents need a bigger inventory in the case of meaningful markers because a marker can only be used in very specific circumstances. The question is now in how far this inventory can be further reduced. One possibility is to conventionalize the features of the controller so that markers can be used more broadly. The controller is the unit that determines the features of its dependents (Corbett, 2006). For example, the noun determines the features of the adjective and article. The features of a controller (e.g.



(a)



(b)

Figure 5: A reduction in inventory complexity from 15 to 4 markers (a) is possible thanks to the feature coercion in nouns that are marked conventionally for one of the four possible feature values (b).

gender or number) are initially semantically motivated but if agents start to relax this constraint then they become compatible with fewer markers. This process is known as *coercion*. Both processes happen in the case of human language evolution but we modeled only the second type.

Figure 5 shows the results of our experiments. If agents are allowed to coerce controllers to be compatible with the agreement markers they already have, then the inventory size becomes reduced with 73%, which means that fewer marker constructions need to be consulted in production and a smaller inventory needs to be learned. Agents need to use again a lateral inhibition dynamics to settle on which conventionalization to adopt. Figure 5b shows the winner-take-all dynamics for a single controller.

4. Conclusion

The theory of cultural language evolution through linguistic selection (Mufwene, 2008; Steels, 2012a) argues that language users are able to come up with different strategies for building their language systems but that they will tend to prefer those strategies that help them construct a better language system. 'Better' means first of all that it satisfies their communicative needs, i.e. that there is enough expressive power to convey the meanings that need to be expressed. Second, it means that language complexity is damped as much as possible along the different dimensions discussed here: inventory complexity, form complexity, processing complexity, learning complexity and population-level complexity (i.e. language variation). Damping complexity is needed to keep the language viable and easier to transmit culturally. This paper did not discuss how new strategies get formed or how agents choose themselves between different strategies based on assessment criteria that they potentially compute themselves. Steps in this direction are shown in other papers (see for example (Bleys and Steels, 2001) and (van Trijp, 2013)). Instead we focused on giving concrete examples of agent-based models to demonstrate how the judicious choice of strategies helps to dampen complexity along various dimensions.

5. Acknowledgements

The authors acknowledge the following financial support: L.S. received an ICREA fellowship to work at the the Institut de Biologia Evolutiva (UPF-CSIC) and a Marie Curie Integration grant to finance collaboration with the VUB AI Lab. K.B. received a grant from the IWT (Belgium) for her PhD research at the VUB AI Lab in Brussels.

References

K. Beuls and L. Steels. Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLoS ONE*, 8(3):e58960, 03 2013. doi: 10.1371/journal.pone.0058960.

J. Bleys and L. Steels. Linguistic selection of language strategies. *Advances in AI - LNCS*, 5778:150–157, 2001.

J.L. Deneubourg Camazine S., N. Franks, J. Sneyd, E. Bonabeau, and G. Theraulaz. *Self-Organization in Biological Systems.* Princeton University Press, Princeton, 2001.

G. Corbett. *Agreement*. Cambridge, Cambridge Textbooks in Linguistics, 2006.

L Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto. Nonequilibrium dynamics of language games on complex networks. *Physical Review E*, 74(3):036105, 2006.

B. de Vylder and K. Tuyls. How to reach linguistic consensus: A proof of convergence for the naming game. *Journal of Theoretical Biology*, 242(4):818–831, 2006.

E. Fuss. *The Rise of Agreement: A Formal Approach to the Syntax And Grammaticalization of Verbal Inflection.* John Benjamins Pub. Co., 2005.

B. Galantucci. An experimental study of the emergence of human communication systems. *Cognitive Science*, 29(5):737–767, 2005.

T. Givón. Topic, pronoun and grammatical agreement. In Charles N. Li, editor, *Subject and topic*, pages 149–188. Academic Press, New York, 1976.

J. Hawkins. *Efficiency and Complexity in Grammars*. Oxford University Press, Oxford, 2005.

B. Heine and T. Kuteva. *The Genesis of Grammar*. Oxford University Press, New York, 2007.

C. Lehmann. On the function of agreement. In M. Barlow and C. Ferguson, editors, *Agreement in Natural Language: approaches, theories, descriptions*, chapter 3, pages 55 – 65. Center for the Study of Language and Information, Stanford University, 1988.

V. Loreto, A. Baronchelli, A Mukherjee, A. Puglisi, and F. Tria. Statistical physics of language dynamics. *Journal of Statistical Mechanics.*, P04006, 1997.

S. Mufwene. *Language Evolution: Contact, Competition and Change.* Continuum International Publishing Cy. New York, 2008.

S. Mufwene. Competition and selection in language evolution. *Selection*, 3(1), 2001.

M. Selten and M. Warglien. The emergence of simple languages in an experimental coordination game. *Proc. Natl Acad. Sci. USA*, 104(18):7361–7366, 2007.

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

L. Steels. A self-organizing spatial vocabulary. *Artificial Life*, 2(3):319–332, 1996.

L. Steels. The origins of ontologies and communication conventions in multi- agent systems. *Journal of Agents and Multi-Agent Systems*, 1(2):169–194, 1998.

L. Steels. Modeling the cultural evolution of language. *Physics of Life Reviews*, 8 (4):330–356, 2011.

L. Steels. *Experiments in Cultural Language Evolution.* John Benjamins Pub. Co, Amsterdam, 2012a.

L. Steels. Self-organization and selection in language evolution. In L. Steels, editor, *Experiments in Cultural Language Evolution*, pages 1–37. John Benjamins, Amsterdam, 2012b.

L. Steels and T. Belpaeme. Coordinating perceptually grounded categories through language. a case study for colour. *Behav. Brain Sci.*, 24:469–529, 2005.

R. van Trijp. Linguistic assessment criteria for explaining language change: A case study on syncretism in German definite articles. *Language Dynamics and Change*, 3(1), 2013.

W. Welmers. *African Language Structures.* University of California Press, Berkeley, 1973.