THE EMERGENCE OF COMPOSITIONALITY, HIERARCHY AND RECURSION IN PEER-TO-PEER INTERACTIONS

JOACHIM DE BEULE
Artificial Intelligence Lab, Free University of Brussels,
Brussels, 1050, Belgium
joachim@arti.vub.ac.be

It is argued that compositionality, hierarchy and recursion, generally acknowledged to be universal features of human languages, can be explained as being emergent properties of the complex dynamics governing the establishment and evolution of a language in a population of language users, mainly on an intra-generational time scale, rather than being the result of a genetic selection process leading to a specialized language faculty that imposes those features upon language or than being mainly a cross-generational cultural phenomenon. This claim is supported with results from a computational language game experiment in which a number of autonomous software agents bootstrap a common compositional and recursive language.

1. Introduction

Compositionality, hierarchy and recursion are universal features of language. By allowing the combination of words into hierarchical phrases which can then recursively be combined into larger phrases, these features allow to make infinite use of finite means in language. Therefore, and because they introduce regularities in a language, they also make a language easier to learn. In short, they may increase a language's fitness as well as that of individual language users. The question remains then how they are selected for.

The mechanism explored in this paper focusses on the increased usability aspect. Language is compositional, hierarchical and recursive because it serves a purpose, and if a feature of language is productive and allows for more effective communication, then individual language users will prefer it over less effective means of communication (Croft, 2007).

The effectiveness of an element of language can of course not be isolated from its learnability and the fact that the entire language community should agree upon it. Hence, like in nativism (Hauser, Chomsky, & Fitch, 2002), the capacity for e.g. recursion is assumed, but it need not be language specific, thereby rendering the problematic question of how it could have evolved for language obsolete. Moreover, this capacity need not be part of a universal grammar imposing itself upon language. Instead, it simply needs to be available for language to recruit

(Steels, 2007). Similarly, we do acknowledge that multi-generational mechanisms like iterated learning (Smith, Kirby, & Brighton, 2003) can be shaping forces of language. However, these only act as second order effects on top of the first order dynamics governed by usability considerations.

To support these claims, a number of computational language game experiments were carried out (Steels, 2002) using the framework of Fluid Construction Grammar (De Beule & Steels, 2005). Such an experiment consists of repeatedly picking a random speaker and hearer from a population of agents (simulated language users) and letting them communicate about scenes. After each interaction both agents update their language inventories to improve their communicative skills. Most of the details of the simulations and the results will be discussed in the rest of the paper, for more information the reader is referred to (De Beule, 2007).

2. Experimental Setup

2.1. *Scenes and Topics*

The scenes about which the agents need to communicate would in English be described by sentences like "Tall blond John kicks beautiful Mary": they always involve two participants each fulfilling either the agent or patient role in an event. Both participants may also be further specified by features (like 'tall', 'blond' and 'beautiful' in the example.)

Scenes are presented to the agents in the form of logical conjunctions of predicates, e.g. the example scene would be presented as:

$$\texttt{tall}(x) \wedge \texttt{blond}(x) \wedge \texttt{John}(x) \wedge \texttt{kick}(x,y) \wedge \texttt{beautiful}(y) \wedge \texttt{Mary}(y)$$

The number of different event-, participant- and feature-type predicates was set to (three times) five. However, an arbitrary number of feature predicates may be present in a scene description[a] according to a binomial distribution with average and standard deviation set to one feature predicate per participant.

The speaker agent does not necessarily describe the entire scene to the hearer: possible topics also include both event participants together with zero or more of the features assigned to them in the scene. On average a topic description contains 2.75 predicates. For example, the above scene specifies 14 topic descriptions, including '$\texttt{John}(x)$', '$\texttt{tall}(x) \wedge \texttt{John}(x)$' etc. Note that the latter description specifies that the arguments to the $\texttt{tall}(.)$ and $\texttt{John}(.)$ predicates are equal. Such co-reference relations also need to be expressed. This can be done using a holistic word (i.e. one word covering both predicates at once, including the equality of their arguments) or else with several words plus a number of grammatical constructions specifying an ordering between them, see e.g. Steels (2005).

---

[a] Some of them may be the same as in 'tall tall John'.

Every interaction, a random scene and associated topic are generated and presented to the speaker. The hearer is only presented with the scene, not the topic. Evidently, he does get to see the utterance generated by the speaker for describing the topic, but only after an efficient communication system has been established will the hearer be able to successfully parse it and hence know what the topic was.

2.2. *Language Model*

An agent's lexicon consists of a number of bi-directional word/meaning mappings. The meaning of a word may be any combination of predicates. All agents start-off with empty lexicons.

Whenever a speaker agent needs to verbalize a topic description, he introduces at most one new word covering all predicates at once for which no word is known yet. Different speaker agents may propose different words for the same meaning. Therefore, every word has an associated synonymy score which is updated according to the well-known lateral-inhibition scheme (Steels, 2002).

An utterance is presented to the hearer as a single string, i.e. without word boundaries. He decomposes it into words again according to the entries in his lexicon. He only proceeds when (presumably) at most one word is unknown, otherwise the interaction fails and the speaker decreases the scores of the words used.

Hearer agents do not know the topic so they can not infer the intended meaning of a word from one interaction only. Therefore, every word/meaning mapping also has an associated probability score representing its estimated correctness. These are updated according to the cross-situational learning algorithm as described in (De Beule, De Vylder, & Belpaeme, 2006). In short, this algorithm allows to combine the information about the meaning of words gained in different situations, while at the same time allowing to cope with inconsistencies caused by changes in word meanings.

Agents prefer those word/meaning mappings with maximum associated synonymy times probability scores. The score of a multiple word analysis is determined as the product of the scores of all words involved. Hence, if one holistic word with high score covers the entire topic description then it might be preferred. If however several more atomic words that only together cover the entire topic description have a higher combined score then these will be preferred. Hybrid combinations are also possible.

After lexicon lookup, all predicates in the topic description are covered by a word (speaker side) or all words in the utterance contribute a number of predicates (hearer side.) The orderings among the words in an utterance expres co-reference relations. The way in which a particular word ordering corresponds to argument equalities in the meaning is determined by the grammar and is something the agents need to agree upon. As was the case for words, speaker agents may introduce new rules of grammar as they need them, and hearers will try to adopt them if possible. And just as agents may use and propose both holistic and atomic words, they may also use and propose different types of grammar rules.

Below are schematically shown a number of example rules for combining words covering predicates of the type specified on the right hand side of the rules (`P` stands for `Participant`, `F` for `Feature`, `E` for `Event` and `S` for `Scene`):

```
P(X)            <-  F(X) P(X)                     (1)
P(X)            <-  F(X) P(X) F(X)                (2)
S(X,Y)          <-  P(X) E(X,Y) P(Y)              (3)
S(X,Y)          <-  E(X,Y) P(X) F(Y) P(Y)         (4)
Type-142(X,Y) <-  F(X) E(X,Y)                   (5)
Type-36(X)      <-  F(X) F(X)                     (6)
Type-726(X)     <-  Type-36(X) P(X)               (7)
Type-76(X,Y)    <-  F(X) P(Y)                     (8)
```

For example, the first rule specifies that if a word or phrase covering a meaning of type feature is directly followed by another word or phrase covering a meaning of type participant, then their arguments should be made equal and the result is a phrase of type participant. Hence, each rule introduces hierarchical structure allowing the subsequent application of rules until all co-reference relations (argument equalities) are expressed and all words are fully ordered. The combination of a number of feature type phrases with a phrase of type participant again results in a participant type phrase if they all have identical arguments (rules 1 and 2 but not rule 8.) Only a limited number of type combinations result in simple result types like this (see rules 1-4.) Most combinations result in the creation of new types (e.g. rules 5-7) which can themselves also be used in other rules (rule 7.) Every agent maintains a private grammar and type system. Both the rules (1) and (2) are recursive, but only the first one allows to express an arbitrary number of feature predicates in combination with a participant predicate. Note that agents not only need to agree upon what constituents to take together (what elements should be on the right hand side of the rules), but also upon their order. For example, some agents may propose the SVO-like rule (2), while others may initially prefer the VSO-like rule (4)[b]. Probability (correctness) and preference (synonymy) scores are used both for reaching a consensus and for determining what analysis to prefer, similar to what happens while learning the meaning of words and during lexicon lookup.

3. Results

Figure 1 shows the evolution of the communicative success for different population sizes measured as a running average. From this graph it can be concluded that

---

[b] However, note that rule (4) requires that a feature-type phrase precedes the object participant-type phrase.

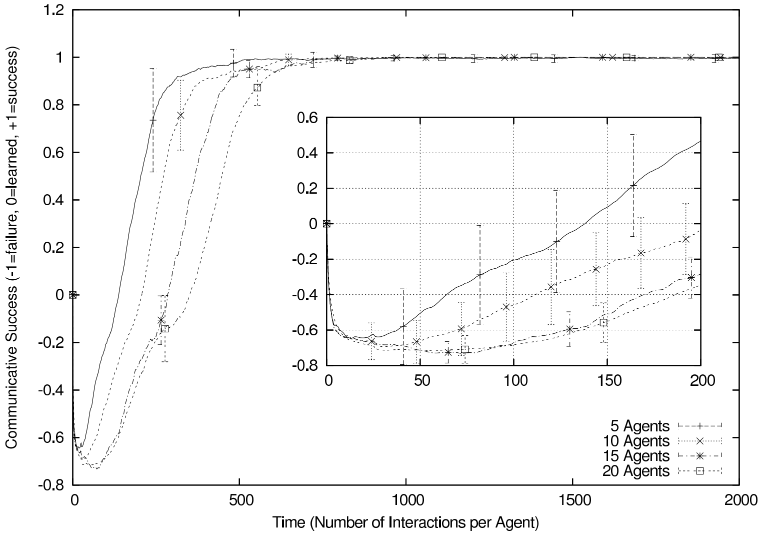

Figure 1. Evolution of the communicative success for different population sizes. (all graphs averaged over 10 independent runs with error bars 1 standard deviation wide.) Time was rescaled such that at time t an agent has had on average $n_a t$ interactions with n_a the population size. The inset shows a detailed portion of the larger graph.

the agents in any case do succeed in evolving a successful communication system.

Figure 2 shows the evolution of the number of predicates in a topic description divided by the number of words in the utterance, measured as a running average. After about 100 interactions per agent, only words are used that have exactly one predicate in their meaning. Put differently: the agents prefer to use compositional language. In contrast to what is the case for communicative success, population size has no influence. This shows that the decision as to go compositional can be made independently from the one about what specific words to use and hence already after a fixed number of interactions per agent rather than after a number proportional to the population size.

Recall that compositional language requires grammar. As it turns out, after about 800 interactions per agent, only rules with result type `Participant` or `Scene` are used (like example rules (1) to (4) but not the others.)[c] Moreover, and again after about 800 interactions per agent, the surviving rules only contain 2 and 3 constituents respectively (i.e. like example rules (1) and (3) but not (2) and (4).) This means that the agents not simply prefer to use compositional and hence grammatical language, but, more specifically, they prefer *recursive* grammar rules

---

[c] Because of space limitations we could not include the relevant graphs here, the interested reader is referred to (De Beule, 2007).

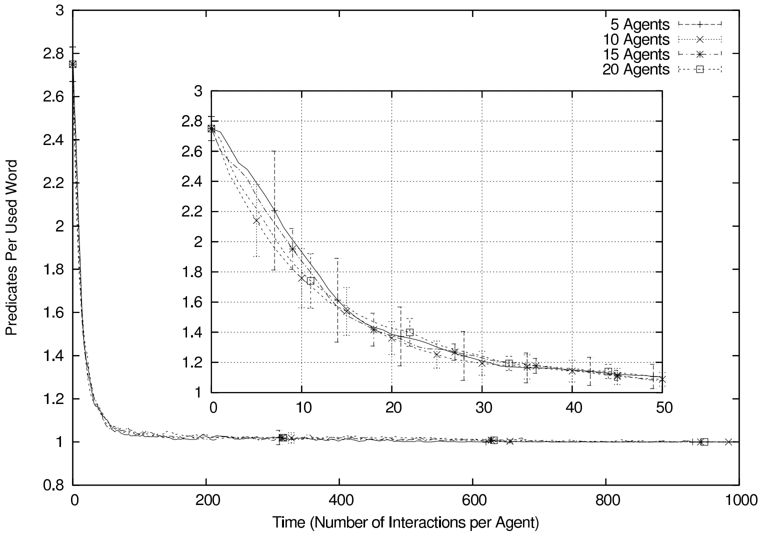

Figure 2. The number of predicates in the topic description divided by the number of words in the utterance. For a completely holistic language this would be 2.75 (the average number of predicates in a topic description.) A fully compositional language would give 1, which is the value to which all graphs converge.

that introduce the *maximum amount of hierarchy.*

4. Discussion and Conclusion

The simulation results confirm that a language can become compositional, hierarchical and recursive simply because language users want to be understood. There is no need to resort to a language faculty dictating these features upon language or to a multi-generational mechanism like iterated learning.

One thing that might appear to be in contradiction with these findings is that natural languages remain partially holistic. Natural meanings are clearly correlated and hierarchically organized. In contrast, the world model considered in the experiments is not. Hence, one cannot expect holistic words to survive because such words simply are of not much use.

If however certain combinations of predicates would appear more frequently in scene descriptions than others, then it *would* be useful to have specific, holistic words for them. This was indeed confirmed in another series of experiments in which the same setup was used as described in this paper except that certain correlations between meaning predicates were introduced. As a result, the emerging languages remained partially holistic (see Figure 3.)

In a third series of experiments the effect of a population turnover was investigated. It should be clear that such a turnover is not required to explain the

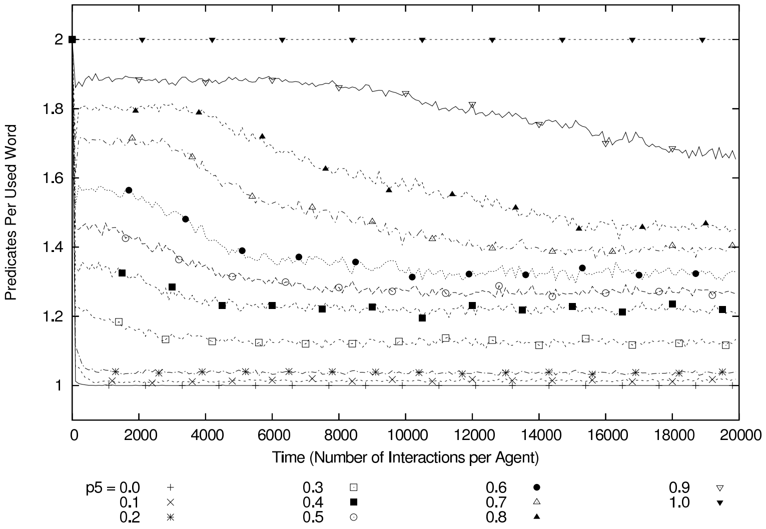

Figure 3. Evolution of the number of predicates covered per used word for different values of a correlation parameter ‘p5’. If p5 equals zero then the experimental setup is identical to the one described in this paper. Increasing values of ‘p5’ correspond to increasing amounts of correlations between otherwise uncorrelated predicates across scene descriptions. For example, if p5>0, then certain participant type predicates will *always* be accompanied by specific feature type predicates (and possibly others.) In topic descriptions they can still occur separately. One can clearly see that larger values of p5 result in on average more predicates per used word, meaning that the agents prefer to use holistic words for frequently occurring combinations of predicates.

emergence of compositionality, hierarchy or recursion. However, since language evolution is a stochastic process, and since iterated learning was shown by others to be a shaping force of language, there are indeed measurable effects. But these are only of second order compared to the first order effects described in this paper, meaning that they are much smaller and only act on a much larger time scale (see Figure 4).

To conclude then, we have shown that the (near) universality of productive features of language like compositionality, hierarchy and recursion can be explained as being an emergent property of the complex dynamics governing the establishment and evolution of a language in a population of generally intelligent interlocutors trying to increase their communicative skills. This happens mainly on an intra-generational time scale. These findings nullify explanations that see natural or cross-generational selection as the main shaping forces of language.

References

Croft, W. (2007). Language structure in its human context: new directions for the language sciences in the twenty-first century. In P. Hogan (Ed.), *Cambridge encyclopedia of the language sciences.* Cambridge: Cambridge University Press.

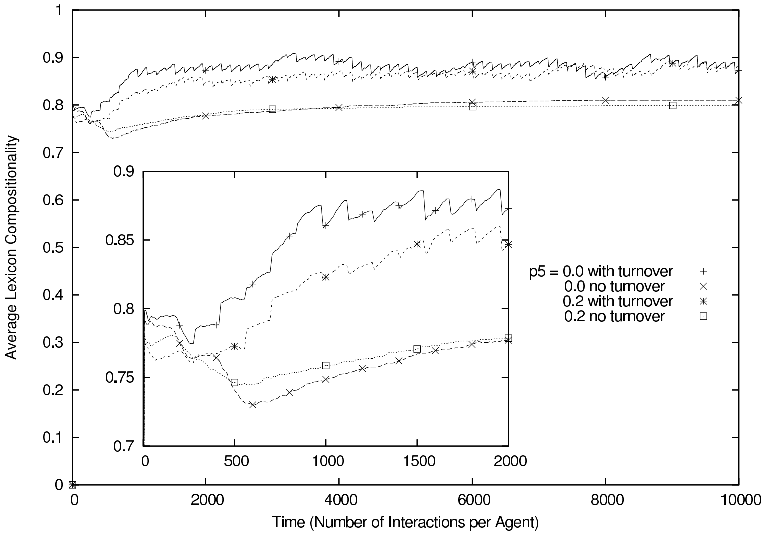

Figure 4. Evolution of the average lexicon compositionality in a fixed population of 5 agents (curves labeled 'no-turnover') and in a population in which after every 600 interactions (i.e. after on average 120 interactions per agent) the oldest agent was replaced by a new one (curves labeled 'with turnover') and for two different experimental settings (see caption of Figure 3 and De Beule (2007) for details.) All curves are based on averaging the inverse of the number of predicates covered by words known by agents, including the words not used anymore but still remembered from earlier phases in the experiment. One can clearly see that, in the case of a turnover, compositionality increases every next epoch, resulting in a larger final degree of compositionality for the curves labeled 'with turnover' compared to the ones labeled 'no turnover'.

De Beule, J. (2007). *Compositionality, hierarchy and recursion in language, a case study in fluid construction grammar.* Unpublished doctoral dissertation, VUB Artificial Intelligence Lab.

De Beule, J., De Vylder, B., & Belpaeme, T. (2006). A cross-situational learning algorithm for damping homonymy in the guessing game. In L. M. R. et al. (Ed.), *Artificial life x* (p. 466-472). MIT Press.

De Beule, J., & Steels, L. (2005). Hierarchy in fluid construction grammar. In U. Furbach (Ed.), *Proceedings of ki-2005* (Vol. 3698, p. 1-15). Berlin: Springer-Verlag.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569-1579.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial Life*, *9*(4), 3710386.

Steels, L. (2002). Grounding symbols through evolutionary language games. In A. Cangelosi & D. Parisi (Eds.), (pp. 211–226). New York, NY, USA: Springer-Verlag New York, Inc.

Steels, L. (2005). What triggers the emergence of grammar? In *Aisb'05: Proceedings of the second international symposium on the emergence and evolution of linguistic communication (eelc'05)* (p. 143-150).

Steels, L. (2007). The recruitment theory of language origins. In C. Lyon, C. L. Nehaniv, & A. Cangelosi (Eds.), *Emergence of communication and language* (p. 129-150). Berlin: Springer.