ON THE EMERGENCE OF COMPOSITIONALITY

JOACHIM DE BEULE
ARTI, Free University of Brussels, Pleinlaan 2, Brussel, 1050, Belgium
joachim@arti.vub.ac.be

BENJAMIN K. BERGEN
Linguistics Department, University of Hawaii, Honolulu, HI 96822, U.S.A.
bergen@hawaii.edu

Compositionality is a hallmark of human language - words and morphemes can be factorially combined to produce a seemingly limitless number of viable strings. This contrasts with non-human communication systems, which for the most part are holistic - encoding a whole message through a single, gestalt form. Why does every human language adopt a compositional strategy? In this paper, we show that compositional language can arise automatically through grounded communication among populations of communicators. The proposed mechanism is the following: if a holistic and a compositional approach are in competition and if both structured (compositional) and atomic meanings need to be communicated, the holistic strategy becomes less successful as it does not recruit already acquired bits of language. We demonstrate the viability of this explanation through computer simulations in which populations of artificial agents perform a communicative task - describing scenes that they have observed. Successful language strategies (that is, those yielding successful transmission of information about a scene) are reinforced while unsuccessful ones are demoted. The simulations show that this reinforcement on the basis of communicative success indeed leads to the dominance of compositional language as long as the fraction of unstructured meaning to be communicated is sufficiently high. Moreover, following Elman (1993), we then show that the same effect can be achieved by, instead of manipulating the world (the fraction of unstructured meaning presented to the agents), letting the agents themselves go through developmental stages. These simulations confirm that simple reinforcement mechanisms applied during communicative interactions can account for the emergence of linguistic compositionality.

1. Introduction

Compositionality is a universal feature of human language. There are varied accounts of where this feature comes from. On the one hand, it could be a requisite characteristic of human language, dictated by an innate universal language capacity endowed upon us by evolution (Chomsky, 1975; Pinker & Bloom, 1990). Alternatively, it could be a cultural innovation, which different language communities composed of generally intelligent humans have consistently converged on because of its tremendous utility. It is the latter explanation that we explore below. We investigate the conditions under which a language community driven only by the success of individual communicative interactions will come to preferentially adopt a globally compositional, rather than a holistic language strategy.

The literature on cultural causes for the emergence of compositional language (Nowak & Krakauer, 1999; Kirby, 2000; Brighton, 2002; Smith, Brighton, & Kirby, 2003) proposes a variety of factors, including iterated learning (IL), learning bottlenecks, expressibility, and the presence of noise in transmission. While all these explanations provide important insights into potential sources of language emergence and change, they leave out three critical considerations.

First, they often address only the emergence of the form and not the meaning facet of compositionality. When linguistic units are brought together to form a composite structure, their forms are assembled (one word might precede the other, for example), but so are their meanings. The grammar of English tells us not only that the sentence "Jack kisses Mary" describes a kissing event involving Jack and Mary, but additionally what roles Jack and Mary play relative to the event. The compositional grammars of Nowak and Krakauer (1999), Brighton (2002) and Smith et al. (2003) are unable to distinguish between "Jack kisses Mary" and "Mary kisses Jack" unless e.g. separate words are used for "Jack" depending on whether he fulfills the agent or patient role in the kissing event, and the same holds for "Mary". Without semantic compositionality, formal compositionality is of questionable utility to language users, and in any case does not approach the expressive power of compositionality in human languages.[a]

Second, in previous work, compositionality emerges at least partially on the basis of experimenter-imposed principles, like an implemented drive to search for generalizations in the learning data and reduce language inventory size (i.e. hypothesize compositional rules). The concern we have with such solutions is that while a trend for decreased inventory size over time (in language learners or populations) might correlate with increased compositionality, implementing this as a causal mechanism effectively causes compositionality to evolve in conformity with the experimenters beliefs about preferred properties of a language system. A more causally satisfying solution would be one in which the language evolves through fundamental principles of communicative interaction, and where changes in inventory size and learnability are byproducts of lower-level causal factors.

Third, IL models typically require many hundreds or thousands of generations for a compositional language to evolve. Moreover, they do not explain how the resulting language could become shared among the members of a larger commu-

[a]It should be noted that this criticism does not apply to Kirby (2000) who simulates a population of agents negotiating about how to express 5 '*event*(?ev,?agent,?patient)' type events.(We adopt a logic-based representation of meaning. Symbols starting with a question mark are variables.) He shows that if agents hypothesize generalizations (basically rules of grammar), then because more general rules are used more often and thus have a better chance of being replicated in the next generation, a shared and compositional language emerges. Despite this quite interesting result, the modeling approach still suffers from two of the concerns mentioned in the text.

nity, as most IL studies consider only one teacher and one learner. We take issue with such models because of the wealth of evidence suggesting that a population of communicators can arrive at a successful, compositional communication system within one or two generations, as evidenced by the emergence of Pidgins and Creole languages and new signed languages, among others.

These properties of previous models all appear to derive from the assumption that the function of language is irrelevant to its form. Because their focus is on how IL can account for the structure and evolution of language, they (often explicitly) disregard linguistic function in their causal models. We present an alternative view on why and how compositional language might emerge, which accounts for both formal and semantic compositionality without relying on endstate-oriented learning mechanisms, and demonstrates convergence on compositional language strategies within a single generation. This view starts from the assumption that the primary function of language is communication. In the experiments we report on, a population of agents is iteratively faced with a communication task in which they talk about a set of observed scenes. The next section spells out the nature of the task and the world in which it is performed. In the following section, a simulation shows how low-level interactive mechanisms - communicative success and a variant of classical conditioning - lead to the emergence of a compositional language when the world obeys certain characteristics. Following (Elman, 1993), we then show how the same effect can be achieved with a wider range of world configurations if instead the agents go through developmental stages.

2. Experimental Setup

Negotiation Model. In our experiments, a speaker and hearer are randomly selected from a population of agents to perform a communication task. They observe a set of scenes, which may differ in terms of their entities or events, the roles the entities play in the events, or combinations of these. The speaker is given one out of the set of events to verbalize. He will use existing language if appropriate, or propose new elements of language if needed, as described below. The hearer decodes the speaker's utterance and tries to identify the topic. If he does so correctly, the game is a success; otherwise it is a failure and the speaker points to the topic. In both cases the agents learn from the outcome.

Other negotiation models could of course be envisioned, and the one used is a radical simplification of the human equivalent. But it is a not unreasonable facsimile of something humans do very frequently both as language learners and language users - namely, observing the world, picking a subset of observations to talk about, and describing that subset using whatever linguistic tools are available. This negotiation framework also importantly allows us to test our main thesis that compositional language can arise automatically in communicative interactions and it does not bias agents towards a holistic or compositional type of solution, only towards a solution that yields communicative success. The same will hold for the language and learning models as explained below.

World Model. Similar to Kirby (2000), the set of potential topics to be verbalized contain partial instantiations of events involving agents or patients of the form '*event*(?ev,?agent,?patient), *person*(?agent)' or '*event*(?ev,?agent,?patient), *person*(?patient)'. Crucially the fact that some entity is playing a particular role in the event, that is, the link (equality) between the ?agent (or ?patient) variable in the event predicate and the corresponding one in the person predicate, is part of the meaning that has to be expressed. If there are N_e different event predicates (e.g. kiss, kick, etc.) and N_p different person predicates (e.g. John, Mary, etc.), this yields $2N_eN_p$ *structured* topics. The *atomic* topics are also included, i.e. one of the $N_e + N_p$ atomic events or people. The fraction of structured topics with which the agents are presented is called the *task complexity*. When it is 0, agents only have to verbalize the atomic topics, and when it is 1, they verbalize only the structured ones. With intermediate values, they eventually get to see them all.

Language Model. The agents are implemented using the Fluid Construction Grammar (FCG) formalism (De Beule & Steels, 2005; Steels, De Beule, & Neubauer, 2005) which is a general unification-based inference engine, designed to support experiments in the self-organization of language. For the current purposes it suffices to say that in the experiments reported on, an agent's language inventory consists of a lexicon and a set of linking constructions which pair word-order with agent and patient role bindings. Agents start with an empty inventory. Whenever a speaker's lexicon does not cover some meaning to be expressed, he creates a new entry, associating a new form with the uncovered meaning. Crucially, the uncovered meaning can be the complete meaning to be expressed or only part of it. For example, if the meaning is '*kiss*(?ev,?agent,?patient), *Mary*(?agent)', and the speaker does not have a word for *kiss* or for *Mary*, he introduces a new word for the entire meaning, which will be holistic. If, however, he already knows a word for *kiss* but not for *Mary*, he will only introduce a new word for *Mary*, which might lead to compositional language. In this case he also needs a linking construction to express the fact that Mary plays the agent role in the kissing event. The speaker has two choices: he can put the predicate either before or after the agent. For successful communication a different word order needs to be used for encoding the agent and patient roles, but whether this solution is found is left to the learning dynamics, based on communicative success. So it may be that a speaker initially uses the same order to encode both the agent and patient cases, or that different agents use different conventions. Newly introduced words or constructions need to be adopted by the hearer. In our model, after pointing a hearer has all the necessary ingredients to adopt any newly introduced words or constructions if at most one word in the utterance is unknown to him.

A strength (a number between 0 and 1) is associated with all lexical and constructional entries. Because all agents introduce new elements of language, an agent's inventory will quickly contain competing entries, e.g. (partial) synonyms and incompatible word order conventions. When verbalizing a topic, a speaker chooses those entries that have the greatest combined strength.

Learning Model. Agents learn by increasing the likelihood that they will reuse successful language strategies, and decreasing the likelihood of reusing unsuccessful ones. Whenever a hearer successfully understands an utterance, he increases the strength s_u of the linguistic elements used according to $s_u \leftarrow \alpha + (1-\alpha)s_u$, where the parameter α is the learning rate, fixed to 0.2. In addition he decreases the strength s_c of all competing elements according to $s_c \leftarrow (1-\alpha)s_c$. This decrease in the strengths of competing entries, a sort of lateral inhibition, is needed to reduce the size of the inventories (i.e. to 'forget' synonyms and reach a coherent language) and introduces a competition between the holistic and compositional strategies. In case of a failed game, both the speaker and the hearer decrease the score of all used entries. This updating scheme can be implemented as an associative memory updated according to the Rescorla-Wagner/Widrow-Hoff learning rule as described in Sutton and Barto (1981). It typically gives rise to the following dynamics. Initially, because agents start with empty inventories, many new elements of language are introduced and the size of language inventories grows. Negotiation then gradually strengthens some of the entries while weakening others and the overall communicative success rises. Entries for which the strength becomes too small are forgotten and the inventory sizes decrease again.

3. The Emergence of Compositionality

Simulations using the framework described above show that a compositional language can only emerge when the task complexity is different from 0 or 1. If it is 0, that is if only the N_e+N_p atomic topics are considered, then the agents never have to express any structured meaning. If it is 1, that is if only the $2N_eN_p$ structured topics are considered, then they never get a chance to introduce words covering only the atomic predications and because the agents do not generalize they again will evolve a stable holistic language. This is also shown in Fig. 1. The simulations also show that a successful communication system only emerges when the complexity is below some threshold (about 0.2 for the simulations shown, or some higher value when either the learning parameter, the population size, or the world size is decreased).[b] These languages are compositional - the bottom figure shows that they require 10 lexical items to communicate the 60 possible topics.

The mechanism at work is the following. Whenever an agent has to verbal-

[b]Communicative success also goes to 1 when the task complexity is 1.0 because in this case the game is equivalent to a standard naming game about $2N_eN_p$ topics in which there is obviously no competition between the holistic and compositional strategies. However for intermediate task complexities this built-in competition seems to prevent the agents to reach a successful communication system.

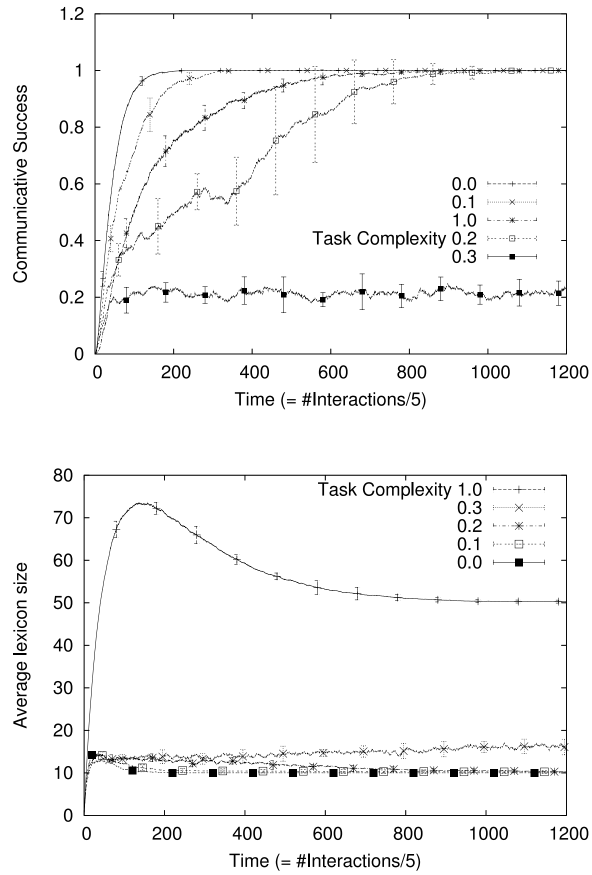

Figure 1. Influence of the task complexity on the evolution of the communicative success and the average lexicon size for several values of the task complexity. The curves were obtained for 5 agents, $N_e = 5$ and $N_p = 5$, averaged over 10 independent runs.

ize structured meaning, his structured and atomic lexical entries compete, and whichever strategy wins will weaken the other. On the one hand, the compositional strategy needs three successful elements of language to be successful itself: two atomic lexical items and one linking construction. In contrast, the holistic approach only needs one. This favors the holistic strategy. On the other hand, this advantage of holistic strategies is counteracted by the fact that a holistic lexical item has fewer opportunities for application, and hence its chances of spreading through the population are smaller, comparable to the replicator dynamics governing Kirby's (2000) IL scheme. The lower the task complexity (as long as it is positive), the more of an advantage the compositional approach displays, since for lower complexities the unstructured (atomic) meanings need to be verbalized relatively more often, allowing the agents to evolve successful simple lexical items, which can then be recruited for communicating structured meanings in a compositional fashion. In sum, then, if the initial task complexity is low enough, the population invariably adopts a successful compositional strategy. (Once successful communication is established, the task complexity is of no further importance.)

One potential drawback of this demonstration of the emergence of compositionality is the constraint that initial task complexity must be low. As explained, this is due to the built-in competition between structured and atomic lexical entries together with the fact that a successful language always requires atomic entries for expressing unstructured meanings. Removing this competition invariably leads to a successful mixed holistic/compositional language, with the degree of compositionality depending on the task complexity. However, as it turns out, there is another configuration that leads to the emergence of a fully compositional language with a broader range of task complexity settings. Following Elman (1993), we can consider whether instead of manipulating the world (the task complexity), we can let the agents themselves go through developmental stages. Elman showed that a gradual increase of attention span or, equivalently, a gradual increase of memory size allowed his neural networks to solve tasks that were unsolvable when starting with a 'full-grown' network. Implementing attention and memory limitations in our agents can be achieved by letting the agents ignore complex situations, i.e. structured topics. To test this, we conducted experiments in which the agents were presented with an equal number of structured and unstructured meanings (i,e, the task complexity was 0.5), but in which a developing learning scheme is adopted. Namely, as long as an agent's communicative success is below 0.7 it ignores all structured topics. Subsequently, as the success increases, gradually more of the structured topics are considered. When 95% success is reached all topics are considered. Not surprisingly, the result is that successful compositional language indeed emerges, even with this higher task complexity. Although more elaborate experiments are needed, the time scale at which this happens seems to be a linear function of the number of rules of language required – i.e. of $N_e + N_p + 2$ in this case: the atomic lexical items plus two constructions.

Such a developmental mechanism has at least two clear correlates in human language use. Perhaps most obviously, children's ability to perceive and conceptualize events is known to increase as part of normal cognitive development. Alternatively, or perhaps in complementary fashion, the development of language about atomic meanings could necessarily precede language about structured meanings; as expertise with language increases, so does the complexity of possible topics. Regardless, the proposed developmental mechanism is a feasible candidate for explaining how humans can arrive at compositional language within one generation, and even more so when a relatively stable simple lexicon is already at hand, which is the case in the development of Creole languages from Pidgins.

4. Conclusion

Compositional language can arise automatically through grounded communication within populations of communicators. This language is compositional both in terms of form and in terms of meaning, and arises within a generation, over the course of hundreds of communicative interactions. Crucially, this is accomplished as the emergent product of simple communication and learning mechanisms. Nothing in the agents' architecture biases them towards compositionality. Instead, the agents are implicitly driven to re-use already established features of language in order to be successful, and are thereby guided towards adopting a compositional strategy.[c] Possibly due to the simplifications made in our agent and learning models, this mechanism only works either when the agents more frequently need to express unstructured meaning or else when they go through developmental stages, which could be explained either cognitively or else in terms of expertise. The emergence of compositionality can be most parsimoniously explained not as an innate universal language capacity endowed upon us by evolution, but through the communicative benefit it brings to individual language users.

References

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*(1).

Chomsky, N. (1975). *Reflections on language.* New York: Pantheon Books.

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

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71-99.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, J. Hurford, & M. Studdert-Kennedy (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form.* Cambridge University Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proc. Nat. Acad. Sci. USA*, *96*, 8028-8033.

Pinker, S., & Bloom, P. (1990). Natural languages and natural selection. *Behavioral and Brain Sciences*, *13*, 707-784.

Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Artificial Life*, *9*(4), 371-386.

Steels, L., De Beule, J., & Neubauer, N. (2005). Linking in fluid construction grammar. In *Proceedings of bnaic-05.* Brussels, Belgium: Royal Flemish Academy for Science and Art.

Sutton, R., & Barto, A. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*(88), 135-170.

Vogt, P. (2005). On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior (Special issue on Evolution and Acquisition of Language)*, *13*(4).

[c] This mechanism might be related to what is called the 'implicit bottleneck' (see e.g. Vogt (2005)) that is introduced because agents are by definition only partially exposed to the emerging language as the language is still under development