Multi-Level Selection in the Emergence of Language Systematicity

Luc Steels(1,2), Remi van Trijp(2), Pieter Wellens(1)

(1) VUB AI Lab - Brussels
(2) Sony Computer Science Laboratory - Paris
steels@arti.vub.ac.be

Abstract. Language can be viewed as a complex adaptive system which is continuously shaped and reshaped by the actions of its users as they try to solve communicative problems. To maintain coherence in the overall system, different language elements (sounds, words, grammatical constructions) compete with each other for global acceptance. This paper examines what happens when a language system uses systematic structure, in the sense that certain meaning-form conventions are themselves parts of larger units. We argue that in this case multi-level selection occurs: at the level of elements (e.g. tense affixes) and at the level of larger units in which these elements are used (e.g. phrases). Achieving and maintaining linguistic coherence in the population under these conditions is non-trivial. This paper shows that it is nevertheless possible when agents take multiple levels into account both for processing meaning-form associations and for consolidating the language inventory after each interaction.

1 Sources of Selection in Language

There is a wide consensus among researchers in language evolution that language must have originated and still continues to evolve because there is a selectionist system underlying it. Despite this consensus, there is however a difference of opinion on how variation and hence the potential for change is caused and what selectionist pressures are operating to retain a particular variation in the language. Basically there are three different types of approaches, based on genetic evolution, cultural transmission, and problem solving respectively.

The genetic evolution models (e.g. [1]) put the selectionist pressure at the level of fitness, which is assumed to be directly related to communicative success. Agents are endowed with an artificial genome that determines how they should communicate: what perceptual primitives they should use for segmenting the world and identifying objects and features, what concepts they can employ for structuring their world, what types of categories are to be used, etc. Potential innovation takes place at the moment this genome is transmitted from parents to children. Because genome copying involves crossover and possibly mutation, variation is inevitable, and some of it will lead to higher or lower success.

Iterated learning models [5] are similar to genetic models, in the sense that variation and hence potential innovation takes place in the transmission of the language system from one generation to the next, but now the language and conceptual system is considered to be culturally coded instead of genetic. Children learn the language from their parents and are then assumed to use it largely unchanged throughout the rest of their life. The learning process necessarily introduces generalisations and variations because of the poverty of stimulus, and hence innovations may enter into the acquired language system. This innovation may re-appear in the data the learners generate for the next generation once they have become adults and thus gets preserved.

The third class of models views the task of building and negotiating a communication system as a kind of problem solving process. Agents try to achieve a communicative goal with maximal success and minimal effort. This problem solving process is definitely not a rational conscious problem solving process but an intuitive one that is seldom accessible to conscious inspection. It is not an individualistic problem solving process either, but a collective one, in which different individuals participate as peers. According to this view a communication system is built up in a step by step fashion driven by needs and failures in communication, and it employs a large battery of strategies and cognitive mechanisms which are not specific to language but appear in many other kinds of cognitive tasks, such as tool design or tool use. Recent experiments on the emergence of communication in human subjects provide good illustrations of these problem solving processes in action [4]. Variation and innovation in problem solving models are common because each individual can invent its own communication system. In fact the main challenge is rather to explain how agreement between individuals and thus a globally shared population language can ever arise.

In the problem solving approach, language becomes a Complex Adaptive System (CAS) in its own right, similar to a complex ecosystem or a complex economy [8]. There are many parallel competitions going on: between synonyms for becoming dominant in expressing a particular meaning, between idiomatic patterns that group a number of words, between different syntactic and semantic categories competing for a role in the grammar, between ways in which a syntactic category is marked, etc. An innovation only survives if it is successful in communication (which could be due to many factors such as the effectiveness of the meanings involved) and if it is also picked up and further propagated by a sufficient number of agents. Often there is no particular reason why one solution is preferred over another one, except that it is more frequent in the population and it wins because of the rich-get-richer dynamics. So we get two types of selectionist forces: functional and frequency-based.

The problem solving/complex adaptive systems approach underlies the many artificial life simulations and robotic experiments we have already carried out in our group (see e.g. [3, 9–12]) and it will be pursued further in the remainder of this paper. Although this approach does not rely on the role of generational transmission in innovation or selection, we nevertheless see clearly a selectionist system arise. As suggested by Croft [2], not grammars (encoded genetically or culturally) but utterances are the 'replicating units' in language. More specifically, we consider the form-meaning associations (for example a word with an associated meaning or a grammatical construction that maps an abstract meaning to a syntactic pattern) as the replicating units. The replication takes place if it is used by a speaker to form an utterance AND if it is processed by the hearer (which may or may not lead to success).

2 Orchestrating Selection through Consolidation

The problem solving/CAS approach requires that the complete communicative cycle is modeled, and not just the transmission from one generation to the next: a specific communicative interaction (a "language game") starts when the speaker sets a specific communicative goal, like drawing attention to an object in the scene, and then conceptualises the world to achieve that goal. He then has to express the conceptualisation based on his own linguistic inventory and produce the resulting utterance. The hearer parses the utterance with his own linguistic inventory and then projects the meaning back into his own experience. If an interpretation was possible, he must choose and perform the action that follows from this interpretation, for example point to the object that the speaker may have had in mind. At the end of this exchange, speaker and hearer can give feedback on success or failure in the game and they may try to fix a failed interaction with additional utterances or gestures.

The highly complex cognitive activities that agents must bring to bear to a complete successful communicative interaction can be grouped into five types of strategies. First of all agents need strategies for setting up a situation in which negotiations can take place to establish a communication system. For example, if we do this with embodied agents (robots) we will have to endow them with ways for setting up a joint attention frame with enough common ground and shared knowledge to guess the meanings that might be expressed by unknown words or constructions. This could be achieved with pointing gestures, eye gaze following, movement towards objects that are going to be the subject of the interaction, etc.

Second, agents need strategies to play the language game itself. They have to make a large number of decisions at all levels of language and conceptualisation. Often there are many alternatives, partly because different conventions are circulating in the population and partly because there is usually more than one way to express similar meanings. These decisions are primarily guided by achieving communicative success.

Agents also need diagnostic strategies for detecting that something is going wrong and for finding out the exact cause. The main indication is of course whether the communication itself did or did not achieve its desired effect. But agents need also more fine-grained diagnostics. For example, a word may have been misunderstood, a perceptual category used by the speaker may have been broader or more restricted compared to that of the hearer, the speaker may have adopted another perspective on the scene than the hearer, etc.

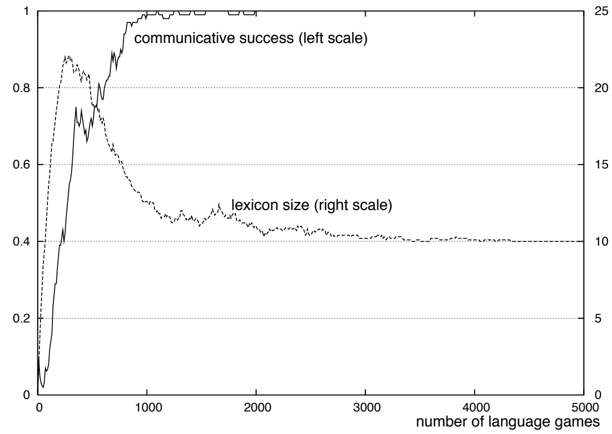

Fig. 1. A population of 5 agents plays a sequence of 5000 naming games, naming 10 objects. We see that the lexicon climbs up at first, reaching a peak of almost 25 after about 200 games. It then settles quickly to an optimum of 10 names for 10 objects thanks to the lateral inhibition dynamics.

Fourth, agents need repair strategies for fixing a problem. For example, agents may introduce a new word or change the definition of a word they have in their lexicon, a hearer may shift a perceptual category slightly to align it with the way that category is used by the speaker, agents may start to mark perspective explicitly, or they may introduce more syntax to tighten combinatorial explosions in the search space or ambiguities in semantic interpretation [13].

Finally, agents need consolidation strategies which are enacted at the end of a game. Based on success or failure, agents have to adapt their inventories in order to become more successful in the future. Each element in an inventory has a particular score and the scores are adjusted to make their usage more or less likely in the future. In this paper we focus on the consolidation strategy, because it is responsible for ensuring that certain form-meaning conventions are selected, independently of functional factors.

These strategies can be illustrated by a simple naming game [7], in which the speaker tries to draw the hearer's attention to an object in the scene using a name. When the speaker does not have a name yet, he introduces a new one. When the hearer does not know the name, he adopts it after additional feedback from the speaker (e.g. a pointing gesture). Because games are played in a distributed population with only local interactions, some pairs may invent/adopt a name and others may invent/adopt another one. Unavoidably different competitors will start to float around in the population. However agents will have the most success and need the least effort if everyone always uses the same name for the same object. This can be achieved if agents keep a score in their local lexicons between objects and names. When they have to choose or interpret a name they use the association with the highest score. When a name is successful, its score is increased and competing names for the same object are decreased (both by speaker and hearer), thus implementing lateral inhibition. When a name is not successful, its score goes down. These strategies create a positive feedback loop so that one name will win the competition (see Figure 1).

The art of setting up language game experiments consists in endowing agents with the right kind of problem solving strategies, such that when these are applied in consecutive language games, the right kind of selectionist process emerges and the global language is driven towards the best or at least a viable state. Note that this process is not implemented directly but is a side effect of the local behaviours of the interacting agents.

3 The Problem of Systematicity

In the case of synonyms, competition takes place between different words for the same meaning, and we have seen that it is settled when agents use the appropriate lateral inhibition strategy. But consider now what happens when different words are grouped. For example, there could be words for actions ("pick-up", "put-down", "give", etc.) and words for objects ("ball", "box", ...) as well as combinations of these in simple patterns ("give ball", "push box", etc.). There will be competition going on among synonymous words (for example between "ball" and "ballon" and "box" and "boîte") but also between the patterns that contain these words (e.g. "pick-up ball" versus "pick-up ballon"). Obviously we want that if "ball" wins the competition at the lexical level, all the patterns that use this choice should become dominant as well. We also want that if new patterns form they should use the "best" words. So we now get selection at two levels: the level of words and the level of combinations of words. A simple lateral inhibition dynamics in which words and combinations compete with each other as if they are on the same level will not do, because there is no guarantee that the winner of the intra-word competition is also the one who wins the intra-pattern competition. For example, it is perfectly possible that the pattern "pick-up ballon" and the pattern "push ball" win their respective competitions, and so there is no guarantee that the meaning of "ball" is always going to be expressed with the same word. Even if words are still used on their own (for example to refer to the objects involved) and hence synonyms get damped (as in Figure 1), there is still the possibility that a word incorporated in a pattern is different from the word that won the competition at the word level.

The next computational experiment shows that this problem indeed occurs in computational simulations (see Figure 2). Agents are now playing games involving an action and objects involved in the action. Sometimes they describe only an object or an action and sometimes they describe the scene as a whole. Agents evolve both words for individual meanings ("give", "ball", etc.) and words for the patterns that combine these meanings (as in "give ball"). Figure 2 shows what happens when agents use the same strategies as in Figure 1. There is an optimal size for the lexicon (10 words), but not for the patterns (which would be 14 two-word patterns and 56 three-word patterns). The reason for this becomes

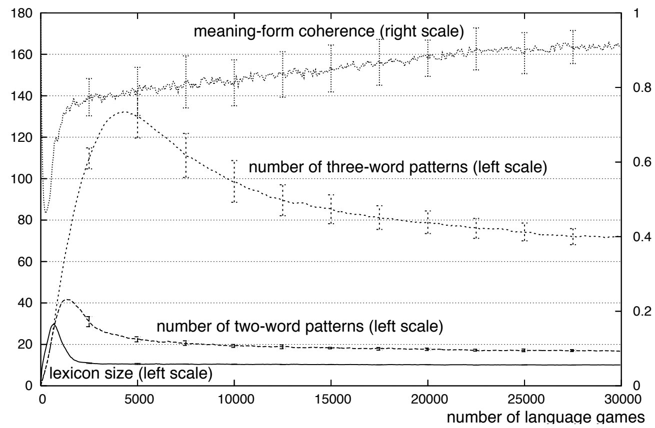

Fig. 2. A population of 10 agents plays a sequence of 30000 games. The graph shows the size of the lexicon and the number of patterns, as well as the meaning-form coherence in the total population which stagnates at around 90%.

apparent when we plot the coherence between meaning-form pairs in the population, measured as the frequency of the most used form for the same meaning. Even though coherence is quite high (around 90%), agents do not reach total coherence and hence have to remember several synonymous, competing patterns in order to maintain communicative success.

This problem of lingering incoherence does not only occur between words and phrases, but also recursively between phrases and more complex phrases. In fact, it has to be dealt with wherever systematicity is desirable in the grammar. For example, a convention for expressing a tense/aspect combination (such as the expression of the present perfective with "have/past-participle", as in "John has seen a lunatic running down the street") should be "productive" across all the situations in which it can be expressed, including in idiomatic phrases.

More generally, this issue of systematicity arises in every selectionist system in which there are elements competing with each other on one level but also undergoing selective competition at another level of larger structures in which they occur. An example from biology concerns the origins of chromosomes. Individual genes started to be combined into larger units (chromosomes). The genes are on the one hand replicators in their own right, undergoing competition, but they are also part of the larger replicating unit of chromosomes [6].

4 Multi-level Selection

We now propose mechanisms that achieve the right kind of multi-level selection. In the first experiment, agents have both a lexicon of associations between meanings and individual words, for example between the predicate [ball] and the word "ball", and a grammar of associations between more complex mean-

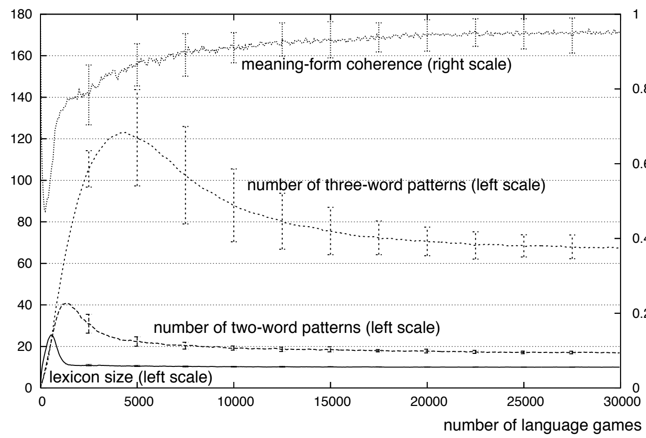

Fig. 3. A population of 10 agents plays a sequence of 30000 games. Here, the agents consider the score of the elements of a pattern in order to choose which pattern to apply. Coherence increases, but doesn't reach 100% yet.

ings and combinations of words, for example between the meaning 'push(o1), ball(o1)' and "push ball". Each of these associations has a score. When agents need to decide how to express a particular meaning they try to cover it with the minimal number of associations that has the highest score. This implies that a pattern, if it exists in the grammar, takes precedence over individual words.

The first step now is to refine the decision criteria by which agents choose which association to use, particularly for patterns. The agents do this in the following steps:

1. First look up all patterns that cover the complex meaning or form to be processed.
2. Next rank these patterns, integrating not only the score of the pattern itself, but also the scores of the individual elements that make up the pattern. This is done in a recursive manner, so that agents can also handle patterns that are themselves part of still larger patterns.
3. If these choices lead to communicative success in the language game, then consolidate the (own) score of the pattern AND the scores of each of the elements in the pattern using the lateral inhibition dynamics discussed earlier.

The ranking of patterns in step 2 ensures that the agents will use those patterns that have not only been used successfully in the past themselves, but also whose elements have been the most successful. The lateral inhibition dynamics of step 3 not only punish conflicting patterns, but also conflicting elements that are used in these patterns.

The effect of this strategy on the semiotic dynamics in the population can be seen in Figure 3. The same conditions hold as in the previous figure: there are 10 agents playing 30000 games. Both the lexicon and grammar size are shown as well as the meaning-form coherence. We see that meaning-form coherence has

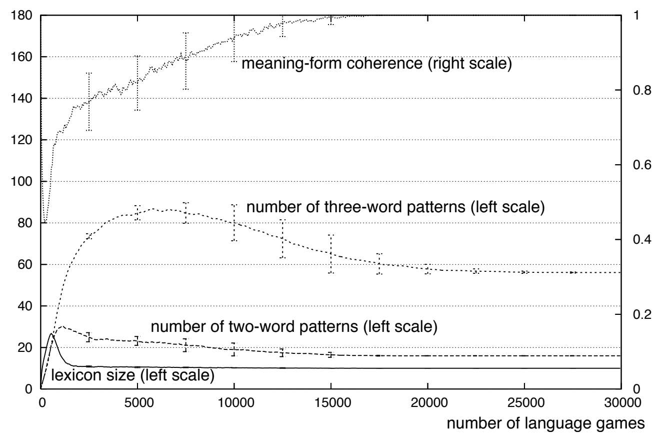

Fig. 4. A population of 10 agents plays a sequence of 30000 games while applying the more sophisticated consolidation strategies. Coherence now reaches 100% and there is greater efficiency in settling the conventions.

improved compared to the previous strategies where patterns and words were competing at the same level, even though it is still not 100%. We also see the agents are slightly more efficient in settling on the grammar, in the sense that 5% fewer patterns circulate in the population before damping sets in.

In a second phase, we further improve results by not only considering the quality of parts to decide on the quality of the whole, but also by investigating in which other patterns this whole is itself a part. For individual words, this means that agents integrate the score of the patterns in which this word is used. Moreover in the consolidation strategy, agents not only update the scores of component parts using lateral inhibition but also the scores of the wholes in which the used association plays a part. Consequently there is now not only a top-down impact (as in Figure 3) from the whole to its parts, but also a bottom-up flow from the parts to the whole. The algorithm can be summarized as follows:

1. First look up which rules cover the meaning or the form to be processed.
2. Next rank these rules, integrating not only the score of the rule itself, but also the scores of all the elements that make up the rule AND the scores of larger rules in which the rule itself is used. This is again done in a recursive manner.
3. If these choices lead to a successful game, consolidate the (own) score of the rule AND the scores of its elements AND the scores of the larger rules of which the applied rule itself is a part.

Figure 4 shows the effect of this more sophisticated strategy for the same conditions as in the previous figures. We see that meaning-form coherence now has further improved to reach total coherence and the number of patterns has

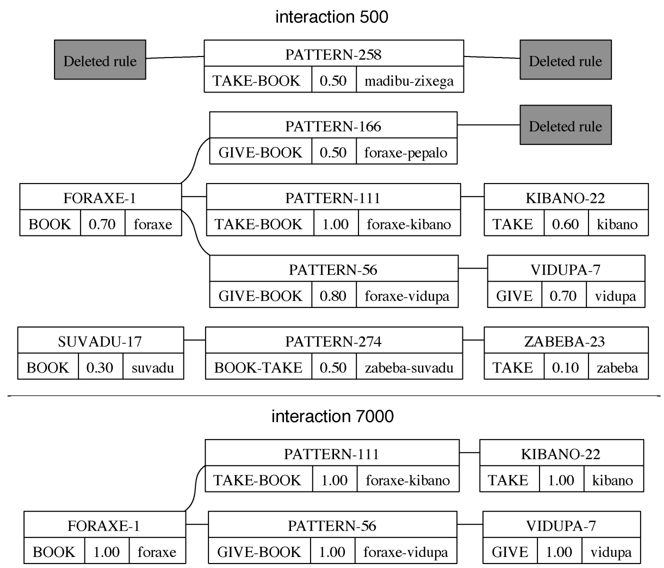

Fig. 5. These two diagrams show the evolution within a single agent's linguistic inventory of the form-meaning pairs covering the meaning "book" and its combinations with "give" and "take" in two-word patterns. Each box contains a rule name (top), the covered meaning (left), a success score (middle) and the form for expressing the meaning (right). The lines indicate which words were used to create which patterns. By considering multiple levels during consolidation, we see that after 7000 interactions the meaning "book" is systematically mapped to the form "foraxe".

reached its optimal size. We also see that there is still greater efficiency in settling on the grammar compared to only top-down impact, in the sense that 25% fewer patterns circulate initially in the population. The effects of this strategy is further exemplified in Figure 5. Here, we see a single agent's network of forms covering the meaning "book" and its combinations into two-word patterns with the actions "give" and "take". The top diagram shows that after 500 language games the agent knows two synonyms for "book" and several competing patterns in which they are used. Some of these patterns even use words which themselves are no longer remembered by the agent. After 7000 games, however, a coherent network has evolved in which "book" is always expressed by the same form.

5 Conclusions

Language can be viewed as a complex adaptive system. Conventions of the language serving the same purpose are in competition with each other for dominance in the population and so in order to reach coherence and hence systematicity in a language, the right kind of selectionist forces must be set up. When all competitors are on the same level (as in the case of competing synonyms) a lateral inhibition dynamics has been shown to be adequate. But when there are different levels in the system because certain elements are used as parts of other structures, we need multi-level selection. We have shown that this can be achieved by endowing the agents with a more sophisticated strategy for choosing which elements they are going to be used in a particular interaction and by applying the lateral inhibition consolidation strategy on different levels.

Acknowledgement. This research was funded (for PW) by a Belgian FWO TM 298 project and (for LS and RvT) by the Future and Emerging Technologies program (IST-FET) of the European Commission under the EU RD contract IST-1940. We thank other members of the VUB AI laboratory, particularly Joachim De Beule, and the SONY Computer Science Laboratory for discussion and implementation of the FCG framework that underlies the experiments.

References

1. E. Briscoe. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2):245–296, 2000.
2. W. Croft. *Explaining Language Change: An Evolutionary Approach*. Longman, Harlow Essex, 2000.
3. J. De Beule and B. K. Bergen. On the emergence of compositionality. In A. Cangelosi, A. Smith, and K. Smith, editors, *The Evolution of Language*, World Scientific, Singapore, 2006.
4. B. Galantucci. An experimental study of the emergence of human communication systems. *Cognitive Science*, 29(5):737–767, 2005.
5. S. Kirby and J. Hurford. The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi, editors, *Simulating the Evolution of Language*, p. 121–148. Springer Verlag, London, 2002.
6. J. Maynard Smith and E. Szathmáry. The origin of chromosomes i. selection for linkage. *Journal of Theoretical Biology*, 163:437–446, 1993.
7. L. Steels. A self-organizing spatial vocabulary. *Artificial Life*, 2(2):319–332, 1996.
8. L. Steels. Language as a complex adaptive system. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. Merelo, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature*, *LNCS 1917*, p. 17–28, Springer, Berlin, 2000.
9. L. Steels and T. Belpaeme. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioural and Brain Sciences*, 28:469–529, 2005.
10. L. Steels and F. Kaplan. Collective learning and semiotic dynamics. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life. Proceedings of ECAL'99*, *LNCS 1674*, p. 679–688, Springer, Berlin, 1999.
11. L. Steels, F. Kaplan, A. McIntyre, and J. Van Looveren. Crucial factors in the origins of word-meaning. In A. Wray, editor, *The Transition to Language*, p. 252–271. Oxford University Press, Oxford, UK, 2002.
12. L. Steels and M. Loetzsch. Perspective Alignment in Spatial Language. In K.R. Coventry, T. Tenbrink and J. Bateman, editors, *Spatial Language and Dialogue*. Oxford University Press, Oxford, UK, 2007.
13. L. Steels and P. Wellens. How Grammar Emerges to Dampen Combinatorial Search in Parsing. In P. Vogt et al., editors, *Symbol Grounding and Beyond, LNCS 4211*, p. 76–88. Springer, Berlin, 2006.