

# Grammaticalization and Semantic Maps: Evidence from Artificial Language Evolution

Remi van Trijp

Sony Computer Science Laboratory Paris, 6 Rue Amyot, 75005 Paris (France)

*Semantic maps have offered linguists an appealing and empirically rooted methodology for describing recurrent structural patterns in language development and the multifunctionality of grammatical categories. Although some researchers argue that semantic maps are universal and given, others provide evidence that there are no fixed or universal maps. This paper takes the position that semantic maps are a useful way to visualize the grammatical evolution of a language (particularly the evolution of semantic structuring) but that this grammatical evolution is a consequence of distributed processes whereby language users shape and reshape their language. So it is a challenge to find out what these processes are and whether they indeed generate the kind of semantic maps observed for human languages. This work takes a design stance towards the question of the emergence of linguistic structure and investigates how grammar can be formed in populations of autonomous artificial “agents” that play “language games” with each other about situations they perceive through a sensori-motor embodiment. The experiments reported here investigate whether semantic maps for case markers could emerge through grammaticalization processes without the need for a universal conceptual space.*

## 1. Introduction

Semantic maps are a powerful research tool for investigating and comparing the grammatical functions of language. Among the many applications of the methodology, a particularly interesting approach is to use semantic maps to identify recurrent grammaticalization pathways. For example, Haspelmath (2003:234) offers a semantic map of typical dative functions with directionality of possible changes. Maps like these make clear and falsifiable predictions about the evolution of grammatical categories across languages.

One of the underlying hypotheses is that semantic maps represent a universal and contiguous conceptual space and that grammaticalization reflects extension or movement of categories along connected regions in this space. Haspelmath (2003:232) writes that if “a semantic map has been tested on a sufficient large number of languages [...] from different parts of the world, we can be reasonably confident that it will indeed turn out to be universal [...].” A similar view has been defended as the *Semantic Maps Connectivity Hypothesis* (Croft, 2001:96). In sum, languages are hypothesized to diverge in terms of grammatical categories, but to share a universal conceptual space.

More recently, the universal status of semantic maps has become a matter of debate. For example, Cysouw (forthcoming) writes that his attempts to find a single and unique semantic map for person marking never led to satisfying results. Instead, he found that several semantic maps are possible depending on the level of analysis and that the traditional semantic maps only offer one particular way of visualizing cross-linguistic variation. Cysouw therefore calls for a different use of semantic maps in which the

number of attestations of a particular function is taken into account. This would lead to a theory of *probable* human languages rather than *possible* human languages.

Unfortunately, non-universalists have not been able yet to propagate a viable alternative for explaining the similarities across languages. For example, Haspelmath (2003:230) argues that without universal conceptual space, “*we would expect languages to differ much more radically from each other than we actually find. Empirical typological work has generally found that similar semantic distinctions are relevant in language after language, independently of genealogical or areal affinities [...].*”

However, ‘expecting’ something is not a sufficient reason for assuming that only universal conceptual space could explain these similarities. The same kind of reasoning has led many linguists to believe in Universal Grammar for decades because languages were considered to be too complex to be learned. However, machine learning techniques have demonstrated how much can be acquired if the learner is granted the right capabilities. Hardwired structures are also difficult to link to the enormous open-endedness and expressivity of human languages, and it is unclear how culture-specific innovations (such as buying and selling, steering an airplane, or sending a robot to Mars) could be mapped onto a universal conceptual space.

In this paper, I will therefore try to contribute to the discussion by proposing and *demonstrating* alternative mechanisms that could explain the ‘universality’ of semantic maps. Instead of looking at natural languages directly, however, I will present experiments on *artificial* language evolution. This methodology has the advantage that it can be used for setting up controlled experiments that investigate features of communication that are typically very hard to grasp for linguists such as the innovation and propagation of linguistic items among multiple language users. For example, most aspects of grammaticalization are only observable to linguists once the change has already occurred so it is very difficult to appreciate the effect of the distributed processes of communication on the structure of language. In artificial language evolution, however, populations of artificial “agents” are used (software entities that model language users) that can be fully inspected by the researcher. Past experiments have already demonstrated how such artificial agents can self-organize a shared ontology and a shared lexicon for objects through general categorization mechanisms and convergence (‘alignment’) strategies (Steels, 1997). These experiments have also been successfully applied (using robots and real-world environments) to the domains of color (Steels and Belpaeme, 2005) and space (Loetzsch, van Trijp and Steels, 2008; Steels and Loetzsch, 2008). In this paper I will present simulations that extend this line of research to the domain of grammar.

More specifically, I will present experiments on the formation of case markers for semantic roles. I will argue that no universal conceptual space is needed for the formation of coherent semantic roles, and that semantic maps could emerge as a side-effect of the need to increase communicative success in locally situated interactions. The coherence of semantic maps is the result of properties of the world and experience on the one hand, and of exploiting analogy for innovation on the other.

The remainder of this paper is structured as follows. The next section offers a brief introduction into the methodology of artificial language evolution and section 3 describes the experimental set-up. Next, the results are reported and discussed. Finally, section 5 ties these results to the study of language typology and grammaticalization and suggests which further steps need to be taken to obtain even more relevant results.

## 2. Artificial language evolution: why should linguists care?

Artificial language evolution is a methodology in which the experimenter tries to find out what is needed in terms of cognitive mechanisms and communicative pressures that would enable agents to autonomously create their own “artificial language” from scratch which features *similar* characteristics as those found in natural languages. A successful implementation does not prove that the natural language phenomena come about in the same way, but at least provides a working hypothesis and hence a possible explanation is offered. The results should therefore be compared to the evidence gathered in other fields such as linguistic typology, psychology and (neuro-)biology. The work in this paper follows a cognitive-functional approach to language in which the methodology typically follows these steps (illustrated here for case marking; also see Steels, 2006, for a more thorough discussion of the methodology):

1. The experimenter picks a topic of interest (e.g. case marking).
2. The experimenter devises a hypothesis about...
  - a. which functional and other external pressures are needed to trigger the development of the topic (e.g. communication about events);
  - b. which cognitive mechanisms are involved in the construction and learning of the phenomenon (e.g. analogical reasoning and convergence strategies).
3. The experimenter then operationalizes the hypothesis...
  - a. by designing an interaction script and a simulated or real world environment;
  - b. by implementing the hypothesized cognitive mechanisms in the form of computational processes and algorithms.

The goal of the experiments is to *demonstrate* that the proposed mechanisms and pressures are indeed necessary and even minimally required for the development of the phenomenon of interest. This can be achieved by comparing series of experimental runs *without* the proposed mechanisms to simulations *with* the proposed mechanisms. In this way the impact of each mechanism on the artificial language can be shown. The methodology therefore does not make any predictions about natural languages, but rather demonstrates the functional and cognitive elements that are necessary for arriving at a certain grammatical stage. Such demonstrations are impossible for natural languages because you cannot “shut down” parts of the human brain, nor can the cognitive apparatus of humans be directly observed (yet).

## 3. Experimental set-up: a problem-solving / usage-based approach

The cognitive-functional approach of this paper is most naturally implemented as a problem-solving model in which the evolution of grammar is driven by communicative needs in language usage. The development of (new) grammar is thus hypothesized to be triggered by the need to increase communicative success and expressiveness and by the need to reduce the cognitive effort required for semantic interpretation. Problem-solving models have also been proposed in various functional and cognitive or usage-based approaches to language. As Ronald Langacker writes:

“[The construction of new symbolic units] is attributed to problem-solving activity on the part of the language user, who brings to bear in this task not only his grasp of linguistic convention, but also his appreciation of the context, his communicative objectives, his esthetic sensibilities, and any aspect of his general knowledge that might prove relevant.” (Langacker, 1991:16)

When language users engage in communicative interactions, it is inevitable that speakers and hearers come across communicative problems. For example, the speaker might not know an adequate or well-entrenched convention yet for expressing a particular meaning. In the experiments reported in this thesis, the agents start without any grammar so they will come across many problems especially in the beginning of the simulations. In order to solve these problems, they are endowed with a rich cognitive apparatus which involves (a) diagnostics for autonomously detecting communicative problems, (b) repair strategies for solving these problems, and (c) convergence strategies for coordinating their linguistic inventories with other agents. In the following subsections, I will explain the cognitive apparatus of the agents, but first I will discuss the communicative pressures and the world environment that cause the communicative problems in the first place.

### ***3.1 Communicative pressures and the world environment***

First of all, it is necessary to find out what functions are typically adopted by case marking systems. As most linguists will agree on, case is a bit like grammar's Swiss army knife in that it can be used for expressing a plethora of grammatical meanings such as event structure (i.e. who's doing what to whom), information structure, determination, spatial and temporal categories, and causal and aspectual relations. Here, I will narrow down the focus of the experiments to the function of marking event structure uniquely and leave out other functional pressures.

A grammar for marking event structure obviously requires communication about events. The artificial agents therefore have to play *description games* with each other in which the speaker has to describe a dynamic event to the hearer. The language game is a success if the hearer agrees with that description and a failure if the hearer disagrees. The experiments make use of data obtained by recording events from a puppet theatre (see Figure 1), which were observed by the agents through two pan-tilt cameras. The event recognition system implements the sensori-motor embodiment of the agents through which they can perceive their world, as described in more detail by Steels and Baillie (2003). Roughly speaking, it uses color recognition and basic visual primitives (such as movement and touching) to detect recurrent visual patterns in the scenes. These patterns are reported as 'macro-events' to the conceptual and linguistic system of the agents together with the 'micro-events' that make up the macro-events.

Each agent in the population can act both as a speaker and as a hearer. Each interaction only involves two agents and cannot be observed by the other agents in the population. All agents are endowed with the same cognitive capabilities and in each experimental run, only one generation of agents is used. All agents are 'adults' which means that their capabilities never change during an experiment so no claims are made about child language acquisition.



Figure 1: The agents engage in description games in which the speaker has to describe a dynamic scene to the hearer. The agents can perceive dynamic real-world scenes through two pan-tilt cameras. This figure shows a puppet walking towards another one.

In order to strictly focus on the development of grammatical structures, the agents are already given a predefined lexicon. This lexicon is a pidgin-like language in which there are words for referring to objects (e.g. *boy* and *block*) and predicates (e.g. *push*, *walk-to*). These words are simple form-meaning mappings and do not have any kind of grammatical specification or categorization, and there are no syntactic or morphological differences for distinguishing word classes. The lexical entries for ‘verbs’ do not contain semantic roles such as ‘agent’ or ‘patient’, but rather their event-specific participant roles as proposed in some current theories of construction grammar (Goldberg, 1995). For example, the verb *push* can involve a *pusher* or something being *pushed*, but it does not contain a predicate frame. Semantic roles such as ‘agent’ and ‘patient’ have to be constructed by the agents themselves during their communicative interactions. There is also no ‘minimal’ lexical entry: no participants are obligatorily expressed and the agents can decide for themselves which part of the event structure they want to profile, which leads to multiple argument realization patterns. For example, *push* can occur in three different patterns: two in which only one participant is explicitly expressed and one in which both participants are profiled. The agents will have to build a grammar that can cope with these multiple patterns of argument structure.

### 3.2 Diagnostics and repair strategies

Language is an inferential coding system in which the intended meaning of the speaker is not always completely covered by the message, as opposed to for example computer programs in which there is always a fixed meaning for a certain form. In other words, the agents have to be intelligent enough to *autonomously* cope with situations in which the utterance of the speaker does not mark every aspect of the intended meaning explicitly. The agents are therefore equipped with a battery of cognitive mechanisms for detecting and repairing communicative problems.

The ability of diagnosing problems therefore presupposes the capacity of interpreting even a partial meaning correctly. In this experiment, the agents have to be able to figure out how objects and events relate to each other. For example, if the speaker has to describe the scene in Figure 1 (in which the boy walks towards the girl) using only lexical entries, he would come up with an utterance such as *walk-to girl boy* in which there is no word order or grammar. If the hearer parses this utterance, he cannot know for

sure who's the walker and who's the destination of the walk-event based on linguistic grounds only. However, if the hearer has witnessed the same scene, he can try to infer the intended meaning by comparing the parsed meaning to his world model. Unless there is too much ambiguity in the context, the hearer can thus successfully infer that the boy was the walker and that the girl was the destination. So communicative success is possible without marking event structure. In natural languages as well, there are many constructions in which event structure is unmarked but in which the context makes clear which reading is intended:

- (1) *the shooting of the hunters*
- (2) *làma nya ánà khù-a* (Lisu; Palmer, 1994:23)  
 tigers TOP dog bite-DECL  
 'Tigers bite dogs.' / 'Dogs bite tigers.'

However, as will become clear when looking at the experimental results in the next section, this requires a lot of cognitive effort from the part of the hearer: for each word in the utterance, he has to figure out how it relates to the other words in the utterance. Moreover, in some contexts there may be closely related events so there might be too much ambiguity for reaching success.

The capacity of inferring the correct meaning can be exploited by the speaker to introduce innovations in the language. Since the speaker wants to reach a certain communicative effect on the hearer (i.e. reach agreement on an event description), he will try to verbalize his intended meaning in such a way as to maximize the chances of communicative success. This requires the speaker to have a model of the hearer to predict the hearer's parsing and interpreting behavior. In these experiments, the speaker will take himself as a model and perform 're-entrance' (Steels, 2003), which can be thought of as some kind of self-monitoring: before transmitting the utterance to the hearer, the speaker will first parse his own utterance himself in order to see what effect it could have on the hearer. If the speaker detects too much ambiguity or too much cognitive effort, he will diagnose a problem. The diagnostic used here is summarized as follows:

**Diagnostic 1:** *If the speaker thinks that the hearer will have to make additional inferences to figure out how the meanings of the individual words are related to each other, a problem of cognitive effort is diagnosed.*

What typically happens in the development of natural language case markers is that a 'light' verb is recruited for solving this problem. For example, Blake (1994:163) gives an example of a serial verb construction in Thai in which the verb *maa* 'come' is used for marking the destination of a fly-event:

- (3) *thân cà bin maa krungthêp* (Thai)  
 he will fly come Bangkok  
 'He will fly to Bangkok.'

The second verb is typically non-finite and takes the same subject as the main verb. Our team is currently working on experiments on such serial verb constructions in order to model this part of the grammaticalization chain. In this experiment, however, the speaker is given the capacity of inventing a new marker (which is specific to the unexpressed relation between the event and the participant role) unless there are other linguistic means of solving the problem that do not require pure invention (see below). For example, if the speaker wants to make the ‘pusher’ explicit in the utterance *boy push* (in which the second participant of the push-event was not profiled), he could invent a marker (let’s say *-bo*) that immediately follows the marked participant:

- (4) *boy -bo push*  
*boy pusher push*  
 ‘The boy pushed (someone / something).’

Since this marker is verb-specific, it cannot be used in other contexts (yet). The repair strategy given here is clearly domain-specific in the sense that the speaker has no choice but to use a marker instead of word order or some other grammatical means. The overall philosophy of our research is however that these domain-specific repair strategies are constructed from more general cognitive mechanisms. For experiments on the recruitment and construction of these mechanisms, see Steels (2007) and Steels and Wellens (2007). The repair strategy can be summarized as follows:

***Repair strategy 1:*** *Unless there are other linguistic means of repairing the problem, and if the problem is not too difficult (i.e. there is only one unexpressed participant role), invent a new verb-specific marker.*

The speaker has also a second repair strategy based on analogy. Since the speaker wants to optimize communicative success, it is better to recycle an existing linguistic item which is probably also known by the hearer. For example, if the speaker has to express the ‘walker’ of a walk-to-event and if he already has the marker *-bo* for marking the ‘pusher’ of a push-event, he will try to reuse this marker in the new situation. Analogical reasoning is used as referee: the speaker will compare the event structure of the walk-to-event to the event structure of the push-event that was used for creating the marker *-bo*. If the participants play analogous roles in both event structures, the marker will be generalized to cover the ‘walker’ as well. In line with usage-based models of language, the extension of a marker is accompanied by an increased productivity (Langacker, 2000). The more participant roles are covered by the same marker, the higher its type frequency and hence the more chance there is that it will be reused again in future interactions. Next to utterances such as (4), the speaker can now produce the following utterance as well in which the meaning of *-bo* has been generalized to the semantic role sem-role-1:

- (5) *boy -bo walk-to*  
*boy sem-role-1 walk-to*  
 ‘The boy walks to (someone / something).’

Innovation through analogy means that grammatical categories expand their usage in a semantically motivated way. This is indeed attested among the world's languages as is clearly visualized by diachronic semantic maps. As opposed to the earlier claims made by Croft (2001) and Haspelmath (2003), however, the extension of semantic roles in this experiment does not require a predefined conceptual space. Instead, the semantically motivated extension is a side-effect of exploiting analogy for increasing communicative success. The repair strategy of analogy can be summarized as follows:

***Repair strategy 2:** If the speaker already knows a marker for a participant role which is analogous to the participant role in the new situation, this marker is generalized and extended into a semantic role. If there are multiple analogies possible, the marker with the highest type frequency is preferred.*

The hearer can learn the innovations of the speaker by using the same cognitive mechanisms for detecting and solving communicative problems. When for example the hearer is faced with the utterance (4) *boy -bo push*, he will not know the form *-bo* yet. The hearer will nevertheless parse the utterance as good as possible and then try to figure out what the speaker might have intended with this innovation. By using the same cognitive mechanisms as used for innovation, the hearer can make an abduction about the meaning of *-bo* and thus learn that it is used as a marker for the 'pusher'.

Similarly for sentence (5), the hearer will notice that *-bo* is used in a different context than before and try to retrieve the analogy intended by the speaker. The hearer will accept the analogy and generalize the marker into a semantic role. If the hearer did not know the marker yet, he will learn it as a specific marker, which can later be generalized during other language games.

Crucial for the claim of this paper is that the algorithm for analogical reasoning (which is described in detail by van Trijp, 2008a:159-162), does not implement an implicit universal conceptual space: event structures consist of recurrent patterns of visual primitives in the world, detected by the event recognition system. Analogies between different event structures are only possible if there are such recurrent patterns. For example, analogies based on movement can only be detected if there are several events that involve the visual primitive of 'movement'. If the agents would live in a world without recurrent patterns, however, the algorithm for analogy would always fail. In other words, there is no pre-wired information about relations between participant roles, but the agents are equipped with adaptive cognitive mechanisms for creating new categories whenever needed.

### **3.3 Convergence strategies**

One of the biggest challenges of multi-agent simulations is to figure out how the agents can align their linguistic inventories with each other: since there are many agents, different solutions for the same problem may get introduced into the population so variation is inevitable. The agents therefore need to have a good convergence strategy that allows them to coordinate their language with each other without the need for central control or without a global overview of the language. This can best be captured by viewing language as a complex adaptive system (Steels, 2000) in which language

becomes an ecosystem in its own right in which many constructions are in competition with each other for becoming the dominant convention in a language. Explaining language change in terms of cultural selection (as opposed to natural selection) is becoming increasingly popular in linguistics as well (Croft, 2000).

Given the fact that this paper mainly aims to demonstrate the power of analogy, I will not go into detail about the various convergence strategies that have been implemented and compared to each other in the experiments. Instead I will restrict myself to the set-up which yielded the most significant results and which shows strong affinity with proposals made in usage-based models of language. A thorough description and explanation can be read in van Trijp (2008a).

Just like speakers of natural languages, the agents will be faced with a lot of variation and competing forms for marking a particular participant role. In order to align their linguistic inventories as much as possible, the agents need a way to decide which variety should be preferred during processing; and they need a way to ‘consolidate’ their communicative experiences in their linguistic inventories.

For processing, the agents will choose the variety that has the highest token frequency. This strategy indirectly favors the more generalized semantic roles over verb-specific markers: semantic roles have a higher type frequency and therefore a wider distribution which results in a higher token frequency. There is thus a rich-gets-richer dynamics for case markers that are more productive than others. At the end of each *successful* language game, the hearer will increase the token frequency score of all the linguistic items that he observed in the utterance of the speaker. At the same time there is a memory decay which decreases the frequency scores of all entries on a regular interval. Infrequent forms can thus be forgotten after some time. To summarize, the processing and consolidation behavior of the agents obey the following rules:

- *Processing*: If an agent knows more than one form for expressing the same meaning, the form with the highest token frequency is preferred.
- *Consolidation*: At the end of a successful language game, the hearer will increase the token frequency of all the observed linguistic items.
- *Decay*: Infrequent forms can be forgotten because of memory decay.

This does not mean that verb-specific markers or ‘smaller’ case markers have no chance of surviving. If a marker can find its ‘semantic niche’ it can avoid being forgotten during memory decay. When discussing the results of the experiment, I will offer a snapshot of competition among case markers.

### ***3.4 Technical issues***

This paper offers a very general overview of the experiment for communicating the results more clearly to linguists who are not familiar yet with the methodology of artificial language evolution. It should be noted, however, that the work described here is built using state-of-the-art techniques in artificial intelligence and computational linguistics. This section therefore gives pointers to other publications for those readers who are interested in the more technical aspects of this work.

First of all, the experiments are implemented using Fluid Construction Grammar (De Beule & Steels, 2005; Steels & De Beule, 2006), a unification-based grammar formalism that has explicitly been designed for supporting experiments on the evolution of grammatical languages. In order to achieve the experiments, it was also necessary to come up with an operationalization of the ‘fusion’ of verb-specific participant roles with the semantic roles of argument structure constructions (Goldberg, 1995). This operationalization had to be flexible and powerful enough to deal with the enormous uncertainty of linguistic conventions that the agents are confronted with when building their language, and it had to be capable of dealing with multiple argument realization. A technical example of the formalism is described by van Trijp (2008b), which also offers the first computational implementation of a construction grammar approach to argument structure that works for both parsing and production.

As for the experiments, more information on the vision system is provided by Steels and Baillie (2003). First results involving two-agent simulations are reported by Steels (2004) and a scale-up to multi-agent populations is handled by van Trijp (2008c). The latter also demonstrates that grammatical languages require a structured network of linguistic items. Instead of traditional top-down inheritance networks, as proposed by most construction grammar theories, an alternative organization is implemented based on ‘multi-level selection’. Steels, van Trijp and Wellens (2007) show that abstract top-down inheritance networks do not suffice for achieving and maintaining systematicity and hence generalization accuracy in a language. The whole case grammar experiment has also been described in detail and with a specific focus on the linguistic relevance of the experiments by van Trijp (2008a).

## **4. Results and discussion**

The above experimental set-up has been implemented in several simulations comparing all relevant parameter settings to each other. Due to space limitations, I will summarize the most important results here. A careful step-by-step overview of the experiments can be found in van Trijp (2008a). All the experiments reported here involve a population of ten agents playing description games that are set in a context containing five events.

### ***4.1 Optimizing communicative success and reducing cognitive effort***

In a first series of simulations, the agents were endowed with the capacity of inferring the speaker’s intended meaning by exploiting the situatedness of the language game. The top graph in Figure 2 shows that the agents can indeed do so. The top line indicates average communicative success which shows that the agents agree with each other in about 70% of the language games. In the other games, failure was due to ambiguity because similar events occurred in the same scene. The bottom line represents cognitive effort which reflects the number of inferences that the hearer has to perform in order to correctly interpret the utterance of the speaker. This score lies between 0 (no inferences needed) and 1 (maximum effort was required). Since the agents had no grammar in this simulation, the hearer always had to make one, two or three inferences (depending on the number of participants involved in the described event). This leads to an average cognitive effort of 60%.

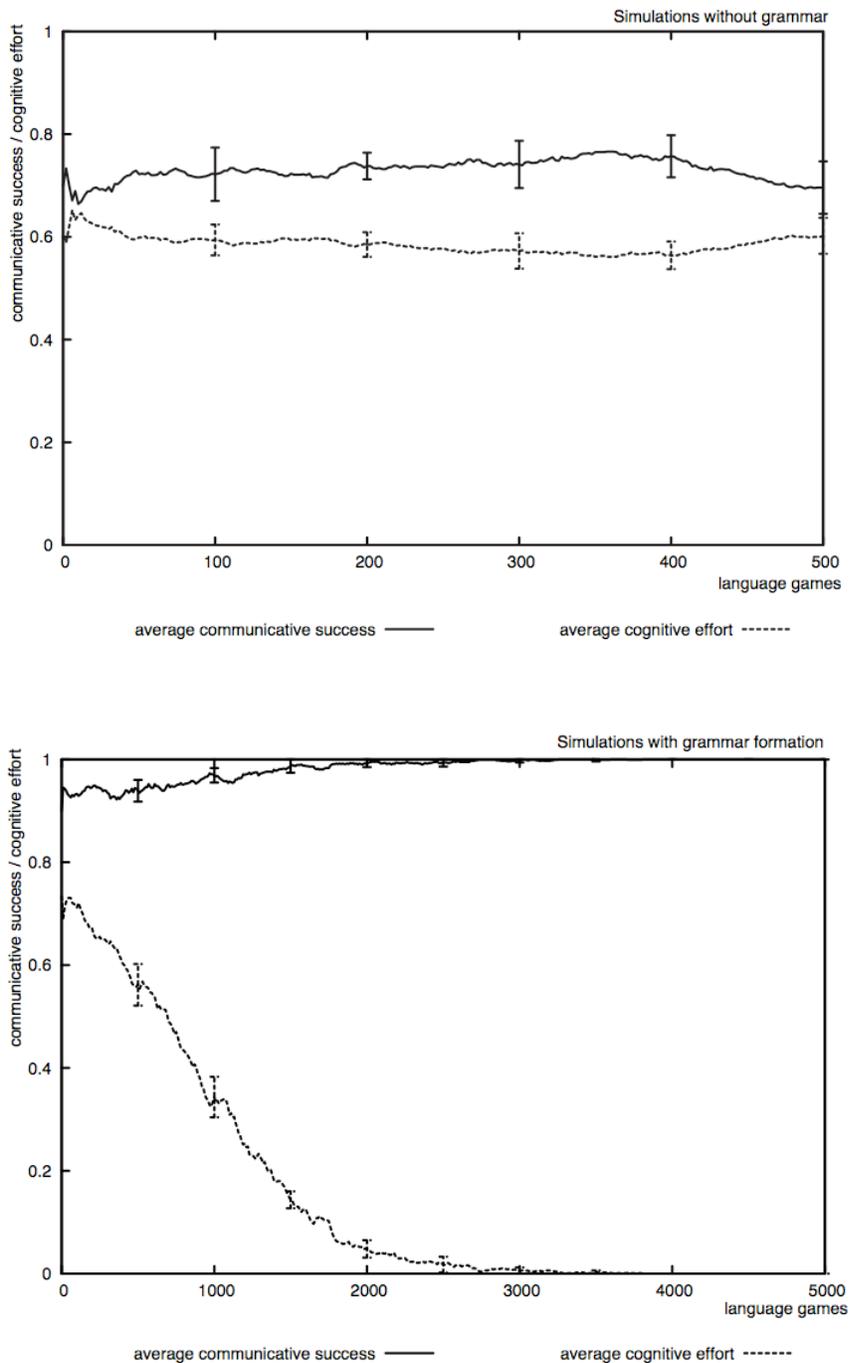


Figure 2: These two graphs show the difference in communicative success and cognitive effort (needed for making inferences) between simulations in which the agents could not make innovations and simulations in which the agents could construct case markers for marking event structure. The top graph shows that the agents can reach a fair amount of communicative success using only a lexical language, but that this requires a lot of cognitive effort on the part of the hearer. The bottom graph shows that case markers improve communication and reduce the cognitive effort needed for interpretation.

The bottom graph shows experiments in which the agents were endowed with the diagnostics, repair strategies and convergence strategies described in the previous section. The graph shows that case markers are indeed useful for communication: communicative success rapidly reaches 100%. This is due to the fact that the markers make the event structure underlying the speaker's utterance explicit, so ambiguity is avoided. At the same time cognitive effort (in terms of additional inferences made by the hearer) drops to zero because the grammar leads the hearer directly to the correct bindings between the participants and their events.

#### 4.2 Generalization as a side-effect of communication

The experiments also show that generalization does not have to be a goal in itself but can emerge as a side-effect of the need to optimize communicative success. By exploiting analogical reasoning, the agents can reuse existing linguistic items and hook new situations to past experiences. This increases the chances that the hearer will be capable of guessing what the speaker intended with his innovation. For the same reason of optimizing communicative success, the speaker will always prefer to reuse frequent and more general markers (those with the highest type frequency): the more frequent a marker is, the higher the chance that the hearer will know it as well. So next to an increase in generalization, also the productivity of linguistic items goes up as the result of communicative pressures.

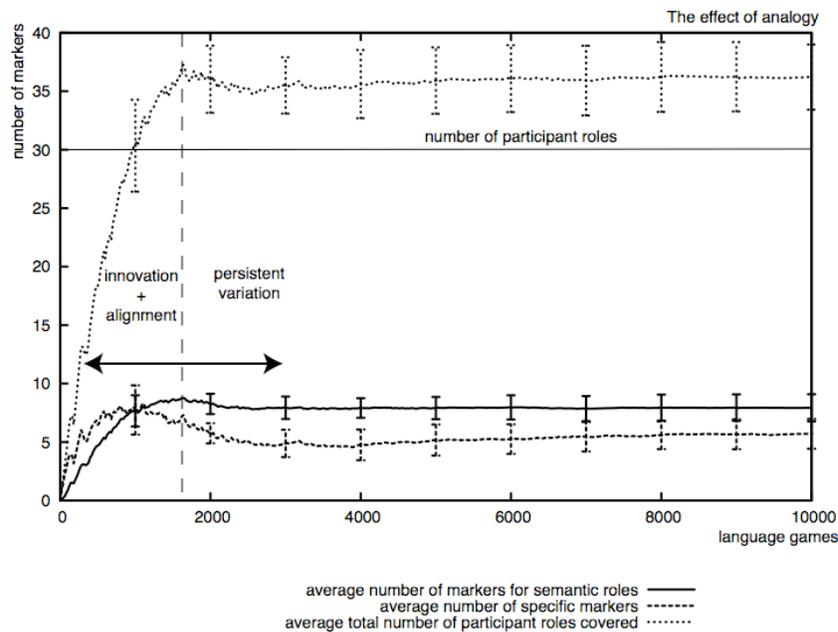


Figure 3: This graph gives an overview of the average number of markers in the simulations. About seven generalized semantic roles have been constructed by the agents and six specific markers survived in their semantic niche.

Figure 3 gives a general impression of the number of markers that are constructed during the experiment and how many participant roles they can cover. The total number of participant roles that need to be covered is 30. We see that the agents start to invent and propagate new markers during the first 2.000 language games after which a period of convergence follows. In the end, about seven generalized semantic roles have become conventionalized units in the language and about six specific markers have found their own ‘communicative niche’ to survive memory decay. We see that there is a bit of overlap between the categories because some variation remains in the population. Overlap of categories is perfectly normal in a language and in fact creates a pool of variation which may cause future language change.

Figure 4 gives a snapshot of the knowledge of a single agent of his language. The graph illustrates the competition among case markers for becoming dominant forms in the language. For each marker, it is indicated how many verb-specific participant roles they can cover. Some markers are clearly more successful than others. For example – *fuitap* rises to eight participant roles within 1.000 language games but then has to give in a couple of participant roles and ends up as the preferred marker for six roles. Other markers die out soon and yet others survive in their semantic niche. The graph clearly shows that analogy causes a continuum from more lexical and specific case markers to more generalized case markers for semantic roles. In natural languages as well, similar continuums have been observed for various grammatical items ranging from more lexical and semantic categories to more grammatical and syntactic constructions.

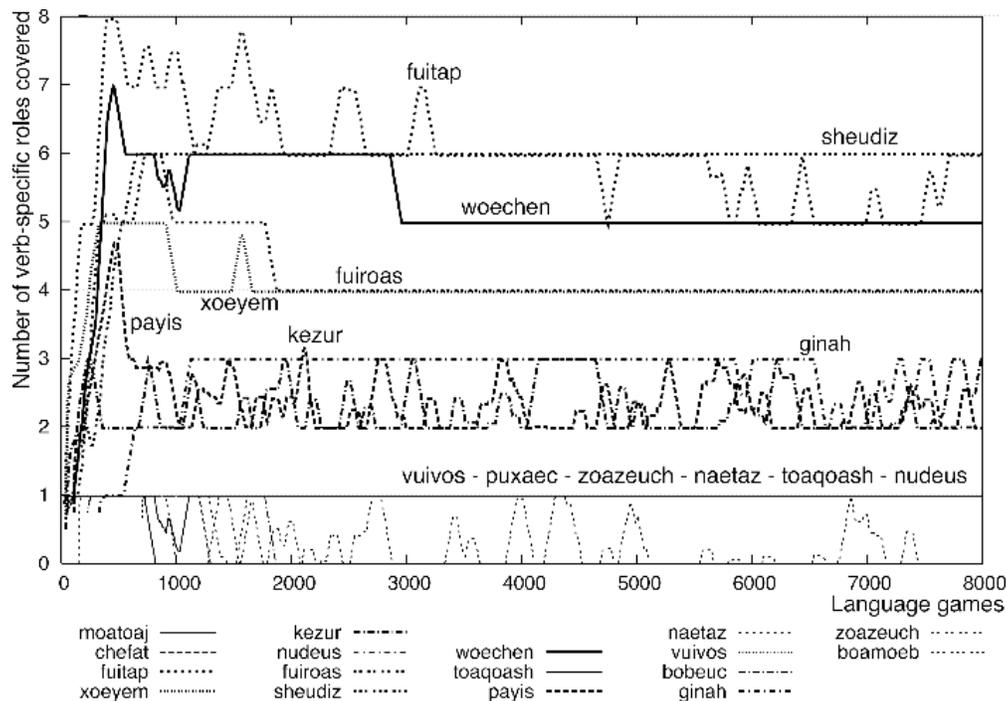


Figure 4: This graph shows the competition among case markers for becoming the dominant form for participant roles as known by a single agent in the population. The graph shows that there is a continuum of more specific and lexical markers to more grammatical cases.

At any moment in the evolution of an artificial language, the researcher can play the role of an ‘artificial language typologist’ and describe the evolved grammar. Here are some example utterances from one of the simulations in which an agent-like semantic role has been formed and covered by the marker *-fuitap*:

- (6) *jack -fuitap walk-to jill -ginah*  
 Jack sem-role-6 walk-to Jill sem-role-26  
 ‘Jack walks to Jill.’
- (7) *touch jill -fuitap house -payis*  
 touch Jill sem-role-6 house sem-role-29  
 ‘Jill touches the house.’
- (8) *house -woechen move-inside boy -fuitap*  
 house sem-role-10 move-inside boy sem-role-6  
 ‘The boy moves inside the house.’

### 4.3 Primitive semantic maps

Finally, this paper is interested in whether the mechanisms and processes described here could explain the formation of semantic maps without the need for a universal conceptual space. Figure 5 illustrates two primitive semantic maps that can be drawn from two different languages that were evolved by the agents in the experiment. The maps suggest that a ‘contiguous’ space is spontaneously constructed as the side-effect of analogy: all categories in the experiment expand their distribution in a radial and gradual way. Similarities across the experiments can thus be explained because the agents apply dynamic categorization mechanisms to recurrent patterns in the world.

As said before, this does not prove that natural languages follow the same strategy, especially given the huge difference in scale between natural language environments and the world implemented in this experiment. The results are however very encouraging and demonstrate the potential power of dynamic categorization mechanisms for explaining systematic patterns across languages. Moreover, no similar models exist yet that demonstrate that a universal conceptual space could yield the same results. Verbal theories often overlook certain flaws of their hypotheses that only come to surface when put to the test of a computational model.

## 5. Conclusions

In this paper I presented experiments on artificial language evolution that investigated how a population of artificial agents could self-organize a case grammar for marking event structure. I argued that the methodology of artificial language evolution can contribute novel evidence to some important debates in linguistics by demonstrating the effect of cognitive mechanisms and functional pressures on the construction of an artificial language. In this paper, I coupled the results of the experiments to the debate on the universality of semantic maps.

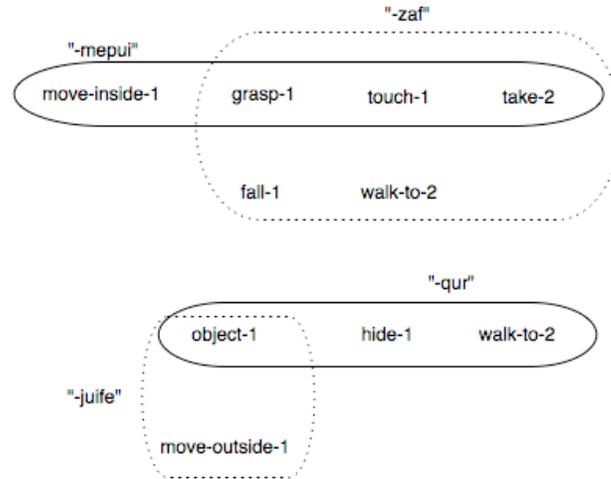


Figure 5: This diagram shows two primitive semantic maps drawn from two different simulations of the experiment. The maps suggest that analogy could lead to similar maps as those observed in natural languages involving contiguous spaces. The continuity of the ‘conceptual space’ arises spontaneously because the extension of case markers typically involves the gradual expansion of a radial category.

I argued that semantic maps do not reflect a pre-wired conceptual space but that they could emerge as a side-effect of locally situated interactions among language users. I substantiated this claim by demonstrating the impact of analogical reasoning on the formation of case grammars for semantic roles: instead of starting from pre-wired relations between participant roles, analogy can create a contiguous semantic map based on recurrent patterns in the world. The experimental results indicated that analogy can explain the continuum of more specific to more grammatical categories, and that extension by analogy happens gradually and in a semantically motivated way. This allowed for a visual representation comparing the categories of two different artificial languages covering continuous areas in ‘conceptual space’. Even though future work is definitely needed for scaling up the experiments to larger worlds, the results are nevertheless very encouraging and offer a viable alternative to universal conceptual space, for which no computational demonstration exists yet.

### Acknowledgement

The experiments reported in this paper were presented at the workshop *Semantic Maps: Methods and Applications* (Paris, 29 September 2007) held adjacent to the seventh meeting of the Association for Linguistic Typology (ALT 7). I wish to thank the organizers of the workshop and the participants for their invaluable feedback, as well as my anonymous reviewer and Michael Cysouw for their insightful comments. I am also greatly indebted to Luc Steels (director of the Artificial Intelligence Laboratory at the Vrije Universiteit Brussel and of the SONY Computer Science Laboratory Paris) and Walter Daelemans (co-director of the CNTS at the University of Antwerp) for their useful comments on this work.

## 6. References

- Blake, B.J. (1994). *Case*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Croft, William (2000). *Explaining Language Change: An Evolutionary Approach*. Harlow Essex: Longman.
- Croft, William (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: OUP.
- Cysouw, Michael (forthcoming). Building semantic maps: the case of person marking. In: Bernard Wälchli & Matti Miestamo (eds.), Berlin: Mouton.
- De Beule, Joachim & Steels, Luc (2005). Hierarchy in Fluid Construction Grammar. In: Furbach, U. (ed.), *Proceedings of the 28th Annual German Conference on AI (KI 2005)*, 1-15. Berlin-Heidelberg: Springer Verlag.
- Goldberg, Adele E. (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: Chicago UP.
- Haspelmath, Martin (2003). The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In: M. Tomasello (ed.), *The new psychology of language*, vol. 2, New York: Erlbaum, 211-243.
- Langacker, Ronald W. (1991). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin: Mouton-De Gruyter.
- Langacker, Ronald W. (2000). A dynamic usage-based model. In: Barlow, M. and Kemmer, S. (eds.), *Usage-Based Models of Language*. 1-63. Chicago: Chicago University Press.
- Loetzsch, Martin and van Trijp, Remi and Steels, Luc (2008). Typological and computational investigations of spatial perspective. In: Wachsmuth, Ipke and Günther Knöblich (eds.), *Modeling Communication with Robots and Virtual Humans*, LNCS 4930. 125-142. Berlin: Springer-Verlag.
- Palmer, F.R. (1994). *Grammatical Roles and Relations*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Steels, Luc (1997). Constructing and Sharing Perceptual Distinctions. In: Van Someren, M. and Wildmer, G. (eds.), *Proceedings of the Ninth European Conference on Machine Learning*. 4-13. Berlin: Springer-Verlag.
- Steels, Luc (2000). Language as a complex adaptive system. In: Schoenauer, M. (ed.), *Proceedings of PPSN VI: Lecture notes in Computer Science*. LNCS, p. 17-26, Berlin: Springer-Verlag.
- Steels, Luc (2003). Language re-entrance and the inner voice. *Journal of Consciousness Studies*, 10(4-5):173-185.
- Steels, Luc (2004). Constructivist Development of Grounded Construction Grammars. In: Daelemans, Walter (ed.), *Proceedings of the Annual Meeting for Computational Linguistics*.
- Steels, Luc (2006). How to Do Experiments in Artificial Language Evolution and Why. In: Cangelosi, Angelo and Smith, Andrew D.M. and Smith, Kenny (eds.), *The Evolution of Language (EVOLANG 6)*. Singapore: World Scientific.

- Steels, Luc (2007). The Recruitment Theory of Language Origins. In: Lyon, C., Nehaniv, C. L. and Cangelosi, A. (eds.), *Emergence of Language and Communication*, p.129-151, Berlin: Springer-Verlag.
- Steels, Luc and Baillie, Jean-Christophe (2003). Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems*, 43, 163-173.
- Steels, Luc and Belpaeme, Tony (2005). Coordinating Perceptually Grounded Categories through Language: A Case Study for Colour. *Behavioral and Brain Sciences*. 28:469-529.
- Steels, Luc and De Beule, J. (2006). Unify and merge in Fluid Construction Grammar. In: Vogt, P. and Sugita, Y. and Tuci, E. and Nehaniv, C. (eds.), *Symbol Grounding and Beyond*, p. 197-223. Berlin: Springer Verlag.
- Steels, Luc and Loetzsch, Martin (2008). Perspective Alignment in Spatial Language. In: Coventry, K.R., Tenbrink, T. and Bateman, J.A. (eds.), *Spatial Language and Dialogue*. Oxford: OUP.
- Steels, Luc, van Trijp Remi and Wellens, Pieter (2007). Multi-level selection in the emergence of language systematicity. In: Almeida e Costa, Fernando, Luis M. Rocha, Ernesto Costa, Inman Harvey (eds.), *Proceedings of the Ninth European Conference on Artificial Life (ECAL 07)*, LNAI (4648), p. 425-434, Berlin: Springer-Verlag.
- Steels, Luc and Wellens, Pieter (2007). Scaffolding language emergence using the Autotelic principle. *IEEE Symposium on Artificial Life 2007*. p. 325-332. Honolulu: IEEE Press.
- van Trijp, Remi (2008a). *Analogy and Multi-Level Selection in the Formation of a Case Grammar. A Case Study in Fluid Construction Grammar*. PhD Thesis. Antwerp: University of Antwerp.
- van Trijp, Remi (2008b). Argumentsstruktur in der Fluid Construction Grammar. In: Fischer, Kerstin and Stefanowitsch, Anatol (eds.). *Konstruktionsgrammatik II: Von der Konstruktion zur Grammatik*. Tübingen: Stauffenburg Verlag. English translation available at: <http://www.csl.sony.fr/downloads/papers/2008/vantrijp-08b.pdf>.
- van Trijp, Remi (2008c). The emergence of semantic roles in Fluid Construction Grammar. In: Smith, Andrew D.M. and Smith, Kenny and Ferrer i Cancho, Ramon (eds.), *The Evolution of Language (EVOLANG 7)*, p. 346-353. Singapore: World Scientific.
- Wellens, Pieter and Loetzsch, Martin and Steels, Luc (2008). Flexible word meaning in embodied agents. *Connection Science*, 20(2), 173-191.