# Can Iterated Learning Explain the Emergence of Case Marking in Language?

Remi van Trijp

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

**Abstract**

This paper compares two prominent approaches in artificial language evolution: Iterated Learning and Social Coordination. More specifically, the paper contrasts experiments in both approaches on how populations of artificial agents can autonomously develop a grammatical case marking system for indicating event structure (i.e. 'who does what to whom'). The comparison demonstrates that only the Social Coordination approach leads to a shared communication system in a multi-agent population. The paper concludes with an analysis and discussion of the results, and argues that Iterated Learning in its current form cannot explain the emergence of more complex natural language-like phenomena.

## 1 Introduction

During the past 15 years, the evolution of language has become an increasingly popular research topic. Many of the advances in the field are achieved through multi-agent modeling and robotic experiments, which allow researchers to create novel artificial languages with similar characteristics as those found in human languages, hence to provide operational explanations for them [11]. This paper compares two influential multi-agent models of *cultural* language evolution, which assume a language-ready brain and which investigate how much linguistic phenomena can be explained without resorting to biological evolution: Iterated Learning [8] and Social Coordination [12]. Iterated Learning Models attribute language evolution to constraints on the transmission of language from one agent to the other, while Social Coordination models hypothesize that language evolves as a response to communicative needs and cognitive effort.

More specifically, this paper contrasts earlier experiments in both models that investigate how a population of autonomous agents can develop a grammatical case system for marking event structure (i.e. 'who did what to whom'), as found in languages such as Latin, German and Turkish. The main finding of the comparison is that Iterated Learning fails to explain the emergence of case, whereas Social Coordination yields promising results. A closer analysis indicates that Iterated Learning does not succeed because it does not require two agents to achieve mutual understanding in their interactions, whereas in Social Coordination models the agents try to align their linguistic inventories with each other in each interaction. The comparison thus strongly suggests that Social Coordination is more suited for achieving advances in the field.

## 2 Case Marking Systems

The experiments compared in this paper investigate the development of early case marking systems. In order to grasp what the experiments are trying to model, it is therefore necessary to take a brief look at the relevant properties of such systems and how they emerge in natural languages.

### 2.1 Function and Properties of Case Systems

In the most prototypical sense, case markers are morphosyntactic inflections that indicate the relations between events and the participants that play a role in those events, as illustrated in example (1) from Turkish:

(1)

| | | | |
|---|---|---|---|
| *Mehmet* | *adam-a* | *elma-lar-ı* | *ver-di* |
| Mehmet.NOM | man-DAT | apple-PL-ACC | give-PAST-3SG |

'Mehmet gave the apples to the man.' [1, p. 1, example 1]

In (1), the case markers make it clear that Mehmet was the *giver* of the apples, because the marker *-a* identified the man as the *receiver*. The accusative marker ı identifies the apples as the *gift*. Marking the roles of event participants has the communicative advantage that utterances can be interpreted without ambiguity and that the hearer does not need to witness the scene in order to infer the intended meaning.

Cross-linguistic research has shown that the case systems of different languages diverge enormously from each other [1]. First of all, not all languages have a case system. Alternative strategies for marking event structure include word order (English), verbal marking (Swahili) or no marking at all (Lisu). Languages that do have a case system, have very different ones – even if the languages are related to each other. Languages vary as to how many cases they have, which types of events they cover, and so on. Case markers are also exploited for other grammatical functions such as number, gender and information structure. This paper however only considers case marking for event structure.

## 2.2 Early Development of Case Systems

Of particular relevance for experiments on the emergence of case is historical data that show where case markers in natural languages come from. A recurrent pattern is that languages recruit an existing lexical item and reuse it as primitive case markers. The following examples from Thai illustrate a typical first step towards the emergence of new case markers in the language [1, p. 163–164]:

(2)

a.

| | | | | |
|---|---|---|---|---|
| *thân* | *cà* | *bin* | *maa* | *krungthêep* |
| he | will | fly | come | Bangkok |

'He will fly to Bangkok.'

b.

| | | | | |
|---|---|---|---|---|
| *thân* | *cà* | *bin* | *càak* | *krungthêep* |
| he | will | fly | leave | Bangkok |

'He will fly from Bangkok.'

Example (2a) features a serial verb construction in which the second verb *maa* ('come') is recruited for indicating that Bangkok is the destination of the fly-event (literally saying 'he will fly and come to Bangkok'), whereas in (2b) a different verb is recruited for marking Bangkok as the city of departure. Both recruited verbs exist as independent lexical items, but here they function as grammatical particles. As well-documented by linguists, such lexical items are at first only recruited for specific contexts, but gradually they become more general through semantic extension [1]. This extension can be visualized using 'semantic maps' that represent the functions of case markers as connected regions in conceptual space [3], showing that case markers start out as semantically coherent categories before they become abstract categories.

# 3 Two Accounts of Cultural Language Evolution

There is a wide consensus that language evolution is caused by a complex interplay between biological and socio-cultural forces. Models of *cultural* language evolution investigate which linguistic phenomena can be explained without resorting to biological evolution. This section compares the foundations of two influential accounts of cultural language evolution: Iterated Learning and Social Coordination.

## 3.1 Iterated Learning

The Iterated Learning Model (ILM) hypothesizes that language evolution is the result of the repeated transmission of utterances from one agent to the next. When there is a learning bottleneck on this transmission (e.g. the number of possible utterances is larger than the number of utterances that can be observed in a lifetime), language adapts to the learning biases of the agents in order to become more learnable. Even though the ILM can be implemented in multiple ways, I will use the name 'ILM' in the remainder of this paper solely to refer to its most widespread set-up.

The ILM implements a population of two agents, in which one agent acts as the adult tutor that needs to verbalize meanings into an utterance, and in which another agent tries to reconstruct the grammar of the adult through induction based on the utterances he observes. Language change is hypothesized to occur as a side-effect of transmission when the learner induces a grammar that is different than that of the adult. After a number of turns, the adult agent is removed from the population and the learner takes his place as the tutor. A new learner is introduced and the cycle repeats itself. The model is illustrated in the left of Figure 1.

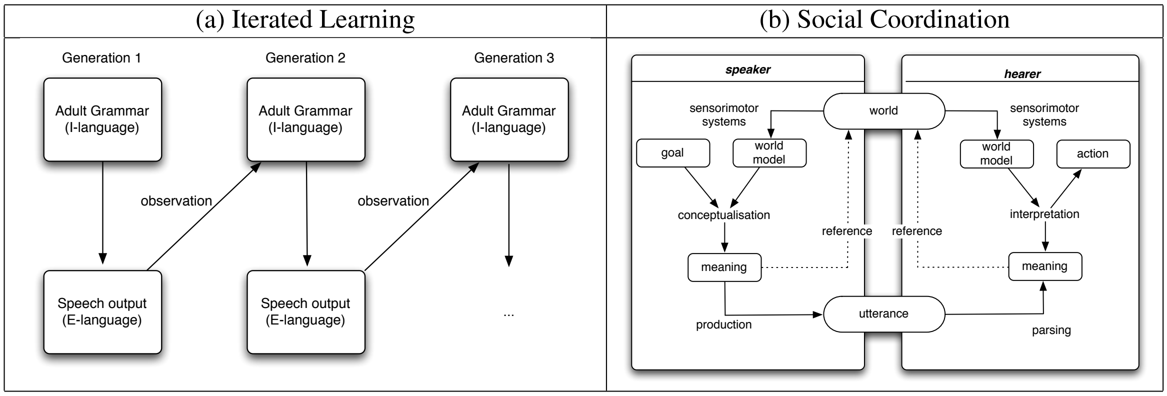


Figure 1: Iterated Learning (left) focuses on the unidirectional transmission from an adult speaker to a child learner, and excludes communicative-functional pressures. Social Coordination (right) models a communicative cycle in which agents can act both as speaker and hearer in peer-to-peer interactions.

The ILM adheres to the 'function independence principle' [2], which states that not all structure in language should be assigned to communicative pressures, hence these pressures are excluded from the model. A crucial modeling choice is therefore meaning transfer: the observing agent always has access to the speaker's intended meaning, which is provided as a "shorthand for the process whereby learners infer the communicative intentions of other individuals by observation of their behavior" [8, p. 374].

A second simplification is that a single agent represents an entire generation, so each generation is assumed to consist of a homogeneous speech population with consistent linguistic behavior. The ILM thus avoids the problem of variation within populations and does not address the question of how new linguistic conventions may propagate in a multi-agent population and become shared in the speech community. However, since the ILM hypothesizes that language evolution is a side-effect of transmission, it subscribes itself to the 'Invisible Hand' theory [5], which assumes that language change occurs when all language users unintentionally make the same decisions.

## 3.2 Social Coordination

Social Coordination models [12, SC] hypothesize that language users continuously shape and reshape their language in locally situated interactions (or *language games*) in order to adapt to their communicative needs and to minimize cognitive effort. Language is obviously still transmitted in these models, but linguistic selection criteria (such as communicative success, cognitive effort and expressivity) play a more important role in the invention and adoption of new linguistic variations than transmission constraints do.

As opposed to Iterated Learning, Social Coordination Models thus consider communicative pressures and cognitive effort as the main driving forces of language evolution. As can be seen in the right of Figure 1, Social Coordination implements a bidirectional communicative cycle in which each agent can act both as the speaker and as the hearer in peer-to-peer interactions. Besides verbalizing and parsing utterances, the communicative cycle also requires the agents to build up a world model, conceptualize meanings (i.e. plan what to say) and interpret them back in their world model. The communicative goal of the speaker and the proper responding action of the hearer are typically given as part of the language game script.

Social Coordination models implement a multi-agent population, which means that variation becomes inevitable as different agents may introduce competing solutions for the same communicative problems. A shared set of communicative conventions can nevertheless be achieved through *self-organization*, which is based on *alignment*. Alignment is the process whereby agents adapt their linguistic inventories after each interaction based on the communicative outcome through reinforcement of successful linguistic items and punishment of competing or unsuccessful ones. Alignment has a long-term impact because it accelerates the self-enforcing loop by which successful linguistic variants are reused in novel situations.

# 4 Experiments on the Emergence of Case

This section summarizes the main results obtained so far on the emergence of case in both models of cultural language evolution. A detailed description of each experiment can be found in [7] for Iterated Learning, and in [14] for Social Coordination. The contribution of this paper is the comparison between both experiments.

## 4.1 Iterated Learning Experiments

The ILM experiments hypothesize that the selective pressure for the emergence of case is free word order. More specifically, agents are hypothesized to retain case-like structures if word order does not offer a reliable cue for retrieving who did what to whom in an event. This hypothesis is investigated through communicative interactions in which an adult agent needs to verbalize a simple two-argument event – such as *loves(john, mary)* – into an utterance, which is then transmitted together with its meaning to a child learner. The meaning space consists of five predicates and five objects, which together amount to 100 possible events. The main objective of these experiments is to expand earlier experiments by Simon Kirby [6] in order to study the formation of a case grammar. Kirby's original experiments investigated how a recursive word-order syntax can emerge as a side-effect of cultural transmission. Despite a learning bias for fully regular grammars, some of the simulations got stuck in a local maximum in which some event participants are verbalized using a different form depending on their occurrence as the actor or the acted-upon in the meaning. For example:

| | | | | | | |
|---|---|---|---|---|---|---|
| | *s/* [P, X, Y] | $\rightarrow$ | 1/X, 2/P, 3/Y | 2/loves | $\rightarrow$ | l, o, v, e, s |
| (3) | 1/john | $\rightarrow$ | j, o, h, n | 3/john | $\rightarrow$ | f, o, o |
| | 1/mary | $\rightarrow$ | m, a, r, y | 3/mary | $\rightarrow$ | b, a, r |

Using the above grammatical rewrite-rules, the meaning *loves(john, mary)* is verbalized as "johnlovesbar", and the meaning *loves(mary, john)* is verbalized as "marylovesfoo". Kirby argues that the two distinct noun categories of these 'suboptimal' grammars can be considered as primitive case systems [6]. The goal of the experiments reported by [7] is to demonstrate which factors *systematically* lead to the emergence of such case-like languages rather than accepting them as local maxima of learning biases.

**Innovation and Adoption.** The adult agent uses a simple set of grammar rules for verbalizing meanings. If the agent has no rule for covering a meaning yet, he invents a random holistic string (i.e. a single form for covering the whole meaning). The experiment also features a 're-order' function that shuffles the words of the speaker's utterances to ensure that word order does not become a reliable cue for parsing. The learner agent is implemented as an eager learner, who tries to make as many abstractions as possible using induction.

**Results and Discussion (1).** The experiments yield three different possible outcomes [7, Ch. 5]. A first type of language (Type A) that emerges is a language that only has one noun category, but different word orders. For example, the emergent language may allow the agents to choose between an S(ubject)O(bject)V(erb) and SVO order. All combinations of word orders are allowed as long as the verb offers a reliable cue for determining which object was the actor and which one was the acted upon, hence SOV and OSV word orders do not occur together because here there is no way of disambiguating them. A second type of language (Type B) is the case-like language with two noun categories as shown in example (3). Type B languages allow any word order combination because the different noun categories act as reliable cues for event structure. A third possible outcome is that the simulation either fails or that the agents never converge on a grammar. The results show that a small amount of reordering (1%) increases the number of Type B languages to about half of the number of simulations, but any further increase in the amount of reordering only leads to increases in the number of failed grammars, with 23.30% of the simulations failing to converge on a grammar if the reordering is set to 10%. Moreover, many of the Type B grammars are inconsistent with respect to which noun category maps onto the 'actor' in the meaning, and which maps onto the 'object', hence they seem to be local maxima rather than case-like languages.

Two conclusions are put forward by the experimenter [7]. First, the results indicate that it is not free word order that selects case, but rather the opposite, that case allows for more variety in word order. The new hypothesis – that case emerges in order to disambiguate utterances – cannot be properly tested because there is no "requirement for agents to *understand* each other, so there *is* no need for disambiguation [7, p. 159]. The experiments therefore suggest that the 'function independence principle' of the ILM is harmful for grammatical structures such as case.

**Results and Discussion (2).** Another problem is that the Type B languages, as they occur in [6, 7] do not correspond to case in natural languages, where case is expressed through markers on the noun (see example 1) rather than through different nouns altogether. Indeed, the ILM grammar inducer "is not capable of *effectively* learning inflectional grammars" [7, p. 206]. In order to scaffold this problem, the semantics of the experiment is enriched through explicit representation of the semantic roles of 'actor' and 'acted-on'. For example, *loves(john, mary)* now becomes *((act, loves), (actor, john), (acted-on, mary))*. Despite this explicit representation of semantic roles, however, the simulations remain problematic in two ways:

- One type of grammar again features two completely distinct noun categories with different words for the same meaning. The inducer fails to recognize inflectional markers for the actor and acted-on.
- A second type *does* have inflectional markers, but their ordering (prefix or suffix) is inconsistent and they also still feature two noun categories, as exemplified in the following two sentences:

(4) *h* *evsn* *kh* *i* *jv* *ib*
actedon john jane actor loves act
'Jane loves John.'

(5) *h* *s* *vs* *i* *jv* *ib*
actedon jane john actor loves act
'John loves Jane.'

## 4.2 Social Coordination Experiments

The Social Coordination experiments hypothesize that case emerges in order to reduce the cognitive load needed for semantic interpretation and to increase expressive power. The experiments discussed here are described in more detail by [14], which is an extension of earlier work by [10]. A similar approach is taken in mathematical modeling by [4]. The goal of the experiments is to demonstrate how a multi-agent population can self-organize a case system for marking event structure, assuming that the population has a shared 'language strategy' (i.e. the same set of operators for innovation, adoption and alignment).

All experiments feature a population size of ten agents that engage in language games about real-world events. In each game, two agents are randomly drawn from the population to act as either the speaker or the hearer. The agents both observe a scene from a puppet theatre in which various actions take place, such as one puppet carrying a block [13]. The speaker then describes one of the actions in the scene and the game is a success if the hearer agrees with that description, or a failure if the hearer disagrees. The agents are equipped with a lexical pidgin language for referring to events and objects, but they have no grammar.

**Conceptualization and Interpretation.** The agents build a world model through sensorimotor processing using an event recognition system called PERACT [13]. The speaker then picks an event to talk about and *conceptualizes* a meaning to express. Meanings are represented using a basic predicate calculus, including an explicit representation of the participant roles of an event. These participant roles are event-specific roles coming from perception (e.g. a 'pusher' and an object being 'pushed'), so no notion of grammar is built in. *Interpretation*, then, is the process whereby the hearer matches a parsed meaning with his own world model. Suppose that the hearer has parsed the following meanings:

```
((boy ?ref-x) (block ?ref-y)
 (push ?ev) (pusher ?ev ?pusher) (pushed ?ev ?pushed))
```

In the above predicates, all symbols with a question mark are variables that can be bound to a referent. The fact that each predicate has its own variable means that the hearer does not know yet how the meanings relate to each other. By looking at the context in interpretation, however, the hearer can *infer* which variables should be made equal. For instance, the variables *?ref-x* and *?pusher* should be made equal if they both refer to the boy in the context. Here, the hearer thus needs to make inferences for both objects in order to retrieve the intended meaning. Inferring variable equalities is experienced by the agents as *cognitive effort*. Case markers can reduce that effort by directly indicating the relations between events and their objects.

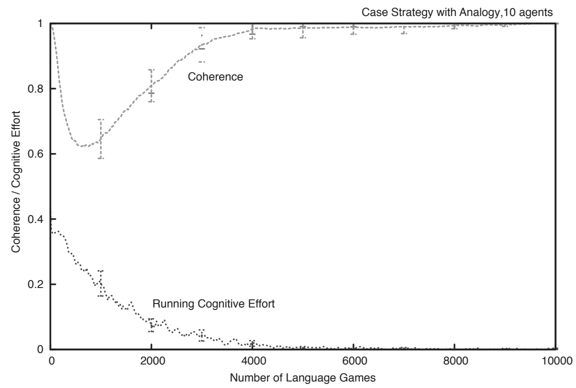


Figure 2: The agents succeed in developing a coherent case system with minimal cognitive effort after 5.000 language games (500 on average per agent).

**Innovation and Adoption.** The experiments implement speaker-based innovation and active learning by the listener. The speaker can predict the cognitive effort needed by the hearer for interpretation through a process called 're-entrance' [9]: before transmitting the utterance to the hearer, the speaker parses his own utterance as if he were the hearer. If he predicts too much effort for the hearer, he will try to avoid this problem through innovation. Similarly, the hearer tries to conceptualize the situation himself based on the partial meanings obtained from parsing in order to retrieve the intended meaning of a novel form. The agents innovate and adopt according to the following strategy:

1. The speaker invents a new case marker or extends an existing one.
   - Diagnostic: The speaker predicts high cognitive effort for the hearer.
   - Repair: If an existing case marker can be semantically extended through analogy, reuse that marker. If no analogous case marker is available, invent a new one.
2. The hearer adopts a new case marker or extends the use of a new one.
   - Diagnostic: An unknown form is observed or a form is used in a novel situation.
   - Repair: If the function of the unknown string can be retrieved from the context, adopt it; or if a known string is used in a novel context, try to extend its semantics to cover the novel situation.

Note that the agents try to reuse case markers as much as possible through analogy rather than inventing new ones. Reuse of existing materials in novel but similar contexts increases the chances that the hearer is able to retrieve the intention of the speaker, hence it increases the chances of communicative success. Each case marker is stored in memory alongside an exemplar of the event that caused its creation. When trying to reuse the case marker through analogy, the agents go into the event structure of the source event obtained from sensorimotor processing (e.g. a push-event can be broken down into various micro-events, such as one visual participant touching another one, then both participants moving together) and map that structure onto the structure of the target event [14].

**Alignment.** A crucial component is the *alignment* performed at the end of each interaction. First, both agents retain innovations in memory if they occurred. Secondly, all constructions in the grammar have a confidence score between 0 and 1, which reflects how confident an agent is that the construction is a conventional meaning-form mapping in his community. New constructions start with a confidence score of 0.5. Only the hearer updates scores at the end of an interaction by increasing the score of all applied constructions in a successful interaction and by decreasing the score of competing constructions through lateral inhibition. The more general case markers (i.e. markers that can be used in many different situations) get rewarded more strongly than specific markers. Updating confidence score impacts the future behavior of the agent, as he always prefers linguistic items with the highest confidence scores.

**Results and Discussion.** The experiments feature a population of 10 agents and a total of 51 'meanings' (i.e. relations between events and their participants) to be covered. Figure 2 shows that the agents succeed in reducing the cognitive effort (Y-axis) they need for interpretation to a minimum after 4.000 language games (X-axis; 400 games on average per agent) by exploiting case, and thus confirms the experiment's main hypothesis. The other line shows coherence (Y-axis), which measures whether all agents use the same case marker for indicating the same participant role. As variation arises in the first 2.000 games, coherence goes down. However, the alignment dynamics restore coherence to its maximum after about 5.000 games (= 500 on average per agent), except for less frequent events, which are settled only later on in the simulations.

The 'semantic map' and graph in Figure 3 provide more details on the type of case systems that emerge. The semantic map (left) shows which participant roles can be covered by two case markers taken from two different populations. As can be seen, the mapping between participant roles and markers is not random, but *semantically motivated* as a side-effect of reusing markers through analogy. The overlap between the two categories shows that analogy is capable of discovering recurrent patterns in the world, which explains why different languages develop similar categories. At the same time, the case categories of both languages provide different categorizations, which also corresponds to findings in natural languages [3]. The graph on the right shows the expressivity of the case markers over time. The Y-axis shows type frequency (the number of different events that a marker can be associated with). There are initially many competitors in the population. However, three markers become dominant and only a few verb-specific cases are able to survive in their semantic niche. The graph shows that the cases form a continuum from specific to general markers, and that their expansion is gradual, as observed in natural languages. Example 6 illustrates that the agents have autonomously developed cases that mark actor- and locative-like relations:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | a. | *Boy-tux* | *move-inside* | *house-pav* | b. | *Jack-my* | *go-away* | *Jill-tux* |
| (6) | | boy+case-12 | move inside | house+case-5 | | Jack+case-10 | go away | Jill+case-12 |
| | | 'The boy entered the house.' | | | | 'Jill walked away from Jack.' | | |

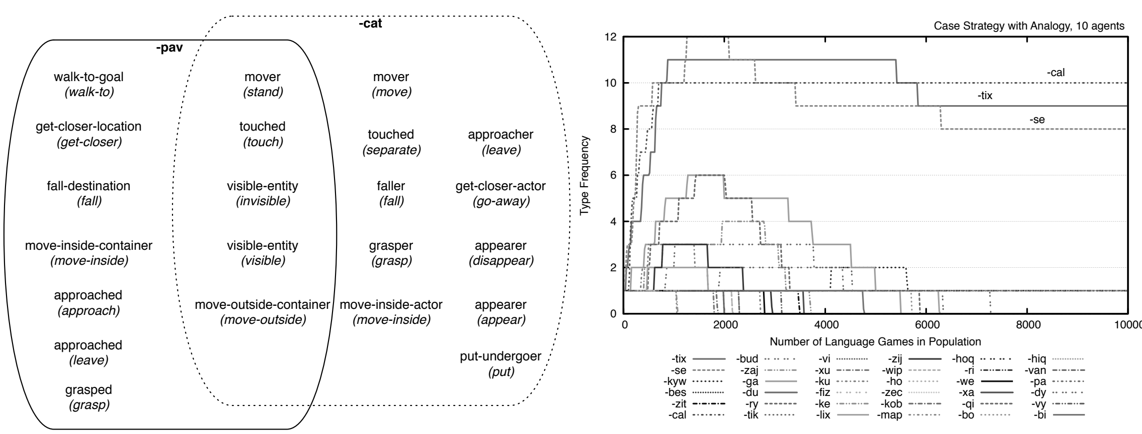


Figure 3: Left: The 'semantic map' compares the participant roles that can be covered by case markers from two different populations. Right: The expressivity of case markers over time.

# 5 Discussion and Conclusion

Even though Iterated Learning and Social Coordination both agree that many aspects of language should be explained in terms of *cultural* evolution, they make different claims about which key processes drive this evolution. The ILM emphasizes constraints on transmission that are independent from communication, whereas Social Coordination focuses on communicative pressures, cognitive effort and social alignment.

The experiments show that Iterated Learning – at least in its most widespread form – does not succeed in evolving a case system; and that the types of languages it evolves show a clear mismatch with widely attested patterns of case development in natural languages, hence it cannot demonstrate that its assumptions work for this domain. The researchers who conducted the ILM experiments conclude that transmission in itself cannot account for case but that agents need to arrive at some mutual understanding. The function independence principle therefore seems to be the wrong modeling simplification to make.

The Social Coordination model, on the other hand, achieved promising results by implementing a usage-based model of language in which agents continuously shape and reshape their language in their local interactions. Even though much work still lies ahead, the model demonstrates that agents may self-organize a case system in order to reduce the cognitive effort needed for interpretation if they are equipped with the right strategy for innovation, adoption and alignment. New case markers emerge for solving concrete problems in communication and gradually extend through analogy, as found in natural languages as well.

The comparison between both models strongly suggests that the specific simplifications of the ILM, particularly the function independence principle and the lack of variation due to the two-agent population, are the wrong ones for addressing complex grammatical phenomena such as case. Indeed, whereas the ILM has so far only been applied to abstract principles of language (such as compositionality and recursion), the Social Coordination Model has already achieved promising results in a wide range of linguistic domains, such as case, agreement, space and tense-aspect [12].

## 6 Acknowledgements

The research reported in this paper was conducted at Sony CSL Paris with additional funding from the EU FP7 ALEAR Project. I thank my colleagues from Paris and the VUB AI-Lab at the University of Brussels; and Luc Steels, director of both labs. All errors of course remain my own.

## References

[1] B.J. Blake. *Case*. Cambridge Textbook in Linguistics. Cambridge University Press, Cambridge, 1994.

[2] H. Brighton, S. Kirby, and K. Smith. Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In M. Tallerman, editor, *Language Origins: Perspectives on Evolution*. OUP, Oxford, 2005.

[3] M. Haspelmath. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello, editor, *The New Psychology of Language*, volume 2, pages 211–242. Lawrence Erlbaum, Mahwah, New Jersey, 2003.

[4] G. Jäger. Evolutionary game theory and typology: A case study. *Language*, 83(1):74–109, 2007.

[5] R. Keller. *On Language Change: The Invisible Hand in Language*. Routledge, London, 1994.

[6] S. Kirby. Syntax out of learning: The cultural evolution of structured communication in a population of induction algorithms. In D. Floreano, J.D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life (ECAL '99)*, pages 694–703, Berlin, 1999. Springer.

[7] J. Moy. *Word Order and Case in Models of Simulated Language Evolution*. PhD thesis, University of York, 2006.

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

[9] L. Steels. Language re-entrance and the 'inner voice'. *Journal of Consciousness Studies*, 10(4-5):173–185, 2003.

[10] L. Steels. Constructivist development of grounded construction grammars. In W. Daelemans, editor, *Proc. 42nd Annual Meeting of the Association for Computational Linguistics*, pages 9–19, 2004.

[11] L. Steels. How to do experiments in artificial language evolution and why. In A. Cangelosi, A. Smith, and K. Smith, editors, *The Evolution of Language (EVOLANG 6)*, London, 2006. World Scientific.

[12] L. Steels, editor. *Experiments in Cultural Language Evolution*. John Benjamins, Amsterdam, 2012.

[13] L. Steels and J.-C. Baillie. Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems*, 43(2-3):163–173, 2003.

[14] R. van Trijp. Grammaticalization and semantic maps: Evidence from artificial language evolution. *Linguistic Discovery*, 8(1):310–326, 2010.