The Evolution of Case Systems for Marking Event Structure

Remi van Trijp

This paper is the author's draft and has now been officially published as:

van Trijp, Remi (2012). The Evolution of Case Systems for Marking Event Structure. In Luc Steels (Ed.), *Experiments in Cultural Language Evolution*, 169–205. Amsterdam: John Benjamins.

Abstract

Case has fascinated linguists for centuries without however revealing its most important secrets. This paper offers operational explanations for case through language game experiments in which autonomous agents describe real-world events to each other. The experiments demonstrate (a) *why* a language may develop a case system, (b) *how* a population can self-organize a case system, and (c) why and how an existing case system may take on new functions in a language.

1. Introduction

There are few linguistic phenomena that have seduced linguists so skillfully as grammatical case has done. Ever since Pāṇini(4th Century BC), case has claimed a central role in linguistic theory and continues to do so today. However, despite centuries worth of research, case has yet to reveal its most important secrets, as can be gleaned from the many recent monographs (Blake, 2001; Butt, 2006), collections (Kulikov et al., 2006; Barðdal & Chelliah, 2009; Malchukov & Spencer, 2009) and projects ('Case and thematic relations', Van Belle & Van Langendonck, 1996; PIONEER, Amberber & de Hoop, 2005; 'Indo-European Case and Argument Structure in a Typological Perspective', Barðdal, 2009). Among the many open questions, the following three are explored in this paper:

1. Why do some languages evolve a case system?
2. How can a population self-organize a case system?
3. Why and how may case systems evolve different functions?

All three challenges have proven to be problematic and even controversial in the literature on case. The first question is answered differently by each linguistic theory – if answered at all. The other two questions have largely been out of reach of linguists due to insufficient historical data, which makes that changes in a language can often only be detected once that change has already propagated and become acceptable in a population (Croft, 1991, p. 34–35).

This paper therefore offers *operational* explanations through experiments in cultural language evolution (see Steels, 2012, and other contributions to this volume). More specifically, this paper proposes the following answers to the above questions and demonstrates their validity through computational models:

1. Case emerges because it has a selective advantage for communication: it reduces the cognitive effort needed for semantic interpretation, while at the same time limiting the required resources for doing so (such as memory space Steels, 2005; Steels et al., 2005).
2. A population may self-organize a case system if all speakers share a language strategy for case (henceforth called: 'case strategy'). Such a strategy comprises components for acquiring, expanding and aligning a case system.
3. Case markers can be repurposed for a different language system if the original selective advantages of a case system have been 'usurped' by more dominant, competing systems in the language.

2. Why Do Some Languages Evolve a Case System?

Among the many functions exhibited by case systems across the world's languages, the most widespread one (and the one investigated here) is to mark *event structure* ('who does what to whom'). Almost every language in the world has developed some grammatical system for marking the relations between events (e.g. a give-event) and the participants that play a role in those events (e.g. a giver, a gift and a receiver). For example, the scene depicted in Figure 1 may be described in English as *the boy is walking to the girl*, using word order and a preposition for distinguishing the walker from the destination. Case languages use nominal markers,

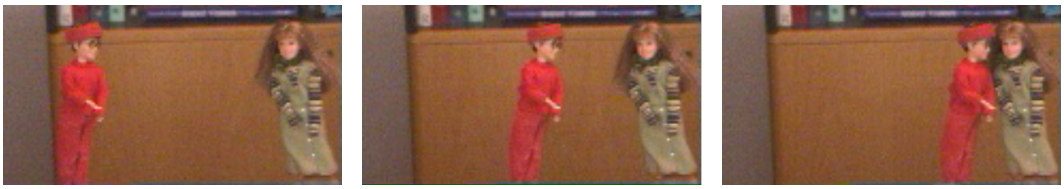

Figure 1. *The agents play language games in which they describe real-world events to each other. In this example scene, one puppet is moving towards another puppet.*

as in the German sentence *Der Junge geht zu dem Mädchen*, in which the definite article *der* assigns nominative case to *Junge* 'boy' and *dem* assigns dative case to *Mädchen* 'girl'.

2.1. The Apparent Complexity of 'Event Structure Grammars'

One striking puzzle in language evolution is why 'event structure grammars' such as case evolve such a degree of complexity if it is perfectly possible to communicate successfully *without* marking the relations between events and their participants (Gil, 2008). For example, some languages such as Lisu (Li & Thompson, 1976) or Riau Indonesian (Gil, 2002) have no or very few grammatical means for indicating event structure. Such languages use lexical items in an associative manner. For instance, Riau Indonesian speakers find it acceptable to say *Cowok lari cewek* 'boy run girl' (David Gil, pers. comm.), a free word order utterance that leaves the correct interpretation up to the listener.

Indeed, language is an inferential coding system (Sperber & Wilson, 1986) in which not all information is explicit in the message, but in which the language user is assumed to be intelligent enough to infer a suitable meaning from the context. So without any clear communicative justification for the degree of grammatical complexity observed in natural languages (Gil, 2008), many explanations for case resort to biological constraints and innate learning biases (Müller, 2002; Moy, 2006) or simply historical accidents (Carstairs-McCarthy, 2004). However, so far none of these theories have been able to demonstrate how a case system may evolve.

2.2. Factual Description Games

I argue that the apparent complexity of case (and other event structure grammars) is the outcome of selective forces that are not always properly appreciated in the literature. In order to identify those selective forces, it is necessary to model situated dialogs (i.e. language games) in which case marking becomes relevant. All experiments reported in this paper involve autonomous agents that play *factual description games* with each other about real-world events (Steels, 2004).

Here is the scenario of a factual description game: assume a population N of agents and a world W consisting of a set of individual objects that participate in dynamic events (such as moving, falling, walking, and so on). Through a camera and an event recognition system called PERACT, the agents are capable of perceiving these events, which are taken from a puppet theatre as shown in Figure 1 (see Steels & Baillie, 2003, for more details on the grounding of events). During each interaction, two agents are randomly selected from the population to act as the speaker and the listener. The speaker selects one of the events from the context and produces a description for it similar to the example sentences from English, German and Swahili used in the previous sections. The listener then parses and interprets the speaker's utterance, and signals agreement if the parsed meanings are valid for the situation that the agents have experienced, or disagreement if there is a conflict.

All the experiments reported in the remainder of this paper use the same world, which consists of 20 scenes recorded using the PERACT system as described by Steels & Baillie (2003). These scenes feature 19 different event types that include 6 monovalent events (such as 'move' or 'appear'), 11 bivalent events (such as 'grasp' or 'approach') and 4 trivalent events (such as 'give' and 'take'), which yields 40 event-specific participant roles. Each scene contains maximally 7 objects (two puppets, two blocks, a house, a table and the ground). In each language game, both the speaker and listener are assumed to have joint attention and to build equivalent situation models in order to rule out interference coming from perceptual differences.

2.3. Reconstruction Experiment: The German Case System

The selective advantage of a language system can be investigated by contrasting the linguistic performance of a population that employs the system to a population that has to make do *without* it. This section compares the performance of a population of agents that are equipped with the German case system to a population that communicates through a lexical Pidgin language. The computational implementation of both languages has been achieved in Fluid Construction Grammar (FCG;

Steels, 2011a), the details of which fall beyond the scope of this paper, but which are described in van Trijp (2011a,b). Both populations start with a lexicon of 22 verbal and 16 nominal constructions for describing the events perceived by the agents. The population equipped with a case system additionally has three argument structure constructions and six 'case markers' (through the German definite articles).

The general architecture of the case language is shown in Figure 2. In production, the agents perceive and conceptualize verb-specific *participant roles* (such as a giver and a gift; top left box) and categorize these roles in terms of more abstract 'semantic roles' (such as Agent and Patient; bottom left box). Argument structure constructions, such as a ditransitive construction, map those semantic roles onto syntactic cases (such as nominative and accusative; bottom right box). Morphosyntactic constructions then provide the surface form of a particular case (e.g. nominative is expressed as *der*, *die* or *das* depending on the syntactic context). In parsing, the same grammatical inventory is used for analyzing and interpreting utterances. Agents that start with a Pidgin language are capable of conceptualizing and interpreting event structure relations (upper left box), but they do not have the grammar for expressing those relations (other boxes).

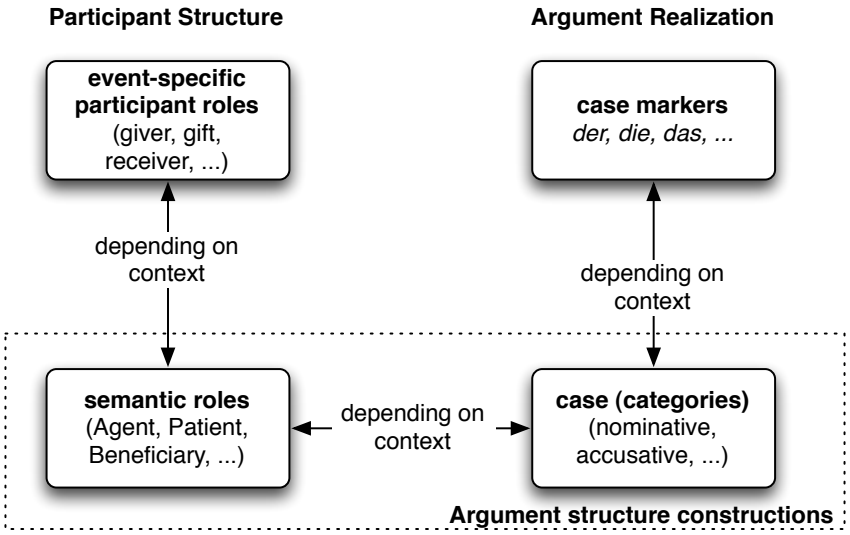

Figure 2. *This 'grammar square' illustrates how the German case system has been 'reconstructed' in Fluid Construction Grammar (van Trijp, 2011a,b).*

Cognitive Effort

In both populations, agents engage in description games with each other. The speakers of Pidgin German, who are deprived of a case system, have no explicit linguistic way of expressing event structure. Instead, they communicate through associative utterances consisting of lexical items. For instance, a Pidgin German agent may describe the scene in Figure 1 as follows:

(1)

Junge	ging	Mädchen.
boy(?x)	*walk(?ev)*	*girl(?y)*
	walker(?ev, ?a)	
	goal(?ev, ?b)	

In example (1), lexical meanings are represented in a first-order predicate calculus under each word. Verbal meanings include a representation for each event-specific participant role involved, which in this example are the 'walker' and the 'goal' of a walk-event. Every symbol that starts with a question mark is a variable that is bound to the referent of that meaning predicate. As can be seen, the meaning predicates do not share any variable with each other because the relations between the words are not marked.

Since the listener has no reliable cues such as word order or case marking for finding out whether the boy was the walker and the girl was the destination (or vice versa, whether the girl was the walker and the boy the destination), he has to *infer* the correct interpretation from the context. Here, the listener makes two inferences: who was the walker and who or what was the destination. The more inferences an agent needs to make, the more *cognitive effort* he requires for interpreting a meaning.

The population with grammar can exploit its case system for marking event structure. In example (2), the case-marked articles *der* 'the.NOM.SG.M' and *dem* 'the.DAT.SG.N' indicate that the boy is the walker and the girl the destination of the walk-event. By marking event structure, the grammar allows the agents to directly interpret the meaning without requiring additional inferences.

(2)

Der Junge	ging	**zu dem** Mädchen.
boy(?x)	*walk(?ev)*	*girl(?y)*
?x = ?a	*walker(?ev, ?a)*	
	goal(?ev, ?b)	*?y = ?b*

Figure 3 shows the running average of cognitive effort needed by the speakers of Pidgin German for correctly interpreting their utterances when playing description games with each other. The graph shows only successful interactions in which the listener was capable of correctly interpreting the speaker's utterance even though the agents have no grammar. The X-axis shows the number of language games on a sequential time scale, that is, at each time step t_s on this scale only two agents are interacting with each other. The Y-axis shows cognitive effort, which is calculated as a Simple Moving Average that computes the mean of the n latest data points for cognitive effort (with $n = 10$ in all experiments reported in this paper). Each data point $c(t_s)$ is the cognitive effort of a language game at a particular time step, which has a value between 0 and 1. This value is obtained by dividing the number of inferences performed by the listener by the maximum amount of possible inferences (which is 3 because the agents maximally express three roles at the same time). The formula for r (running average of cognitive effort) is as follows:

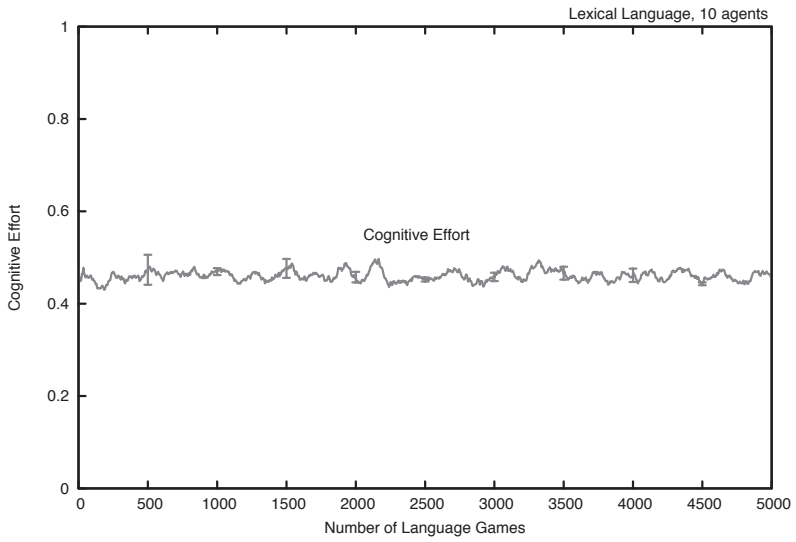

Figure 3. *This graph shows the baseline cognitive effort needed for interpretation by the Pidgin German agents with a population size of* $N = 10$ *on a sequential time scale.*

$$\mathrm{r}(t_s) = \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{c}(t_s - i) \tag{3}$$

The results in Figure 3 represent an average over the values observed in ten experimental runs (henceforth called 'series'). The error bars show standard error. As can be seen in the chart, the speakers of Pidgin German always need to make additional inferences, which leads to an average cognitive effort of about 50%. The population of German speakers, on the other hand, reduce this baseline effort to a minimum of zero in their interactions. Case therefore has a clear selective advantage for communication: it reduces the cognitive effort needed for semantic interpretation.

Expressive Power and Learnability

An alternative way to reduce the cognitive effort for interpretation is to develop idiosyncratic case markers for the relations between events and their participants. Similar to a Naming strategy (Steels, 1995; Steels & Loetzsch, 2012), the agents can invent a word for each new participant role they wish to express. Thus, for a 'walk-event' they would also invent a word for the 'walker' and the 'goal', for a 'give-event' they would invent words for the 'giver', the 'gift' and the 'receiver', and so on. Inventing idiosyncratic cases is not enough, however, because the agents still need a minimal amount of syntax for indicating which markers belong to which words. An idiosyncratic case system is nevertheless 'simpler' than an abstract one in the sense that there is a one-to-one mapping between the meaning and form of an idiosyncratic case as opposed to the polysemous cases found in natural languages.

So what is the selective advantage of a grammatical over an idiosyncratic case strategy? The answer is expressivity and learnability. Figure 4 uses the relation between the number of cases and the number of meanings they are able to express as an indicator for both linguistic pressures. The X-axis represents the number of participant roles that the agents need to express and the Y-axis shows the number of cases that a language needs for expressing those participant roles. As can be seen, the German case system stays stable at four cases despite a growing meaning space and can potentially handle an infinite number of participant roles. For the lexical language that uses names for each new participant role, however, the number of cases is directly proportional to the number of meanings that need to be expressed, hence we see linear growth as the meaning space expands. The results thus indicate

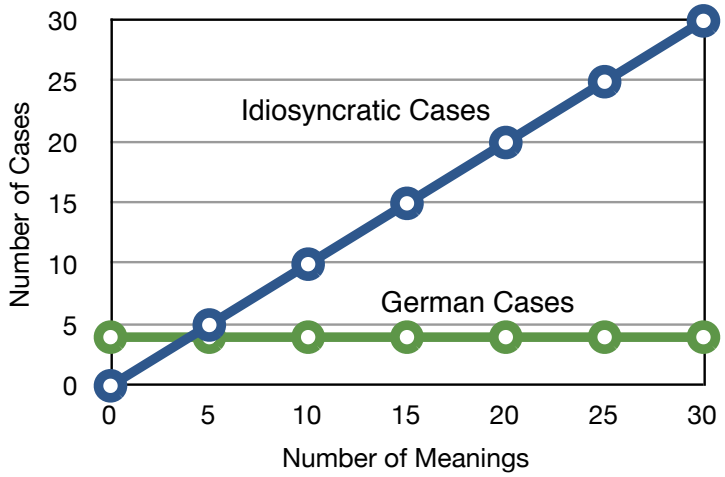

Figure 4. *This chart shows that the German case system can handle a growing number of meanings without increasing its number of cases. A language that uses an Idiosyncratic Case Strategy names all meanings explicitly, however, shows linear growth in its inventory size as the meaning space becomes bigger.*

that a case system is far more economical and requires less memory and is therefore more learnable.

3. How Can a Population Self-Organize a Case System?

The previous section illustrated *why* languages may develop a case system by demonstrating the selective advantages of such a system for communication. Now that the selective pressures are identified, the next step is to show *how* a population can develop a shared case system. Evolving a case system is an extraordinarily difficult challenge that raises many problems, all of which are still a matter of great debate. More specifically, the agents need to:

1. Construct a systematic way for expressing an open-ended meaning space of event structure relations.
2. Acquire and evolve larger argument structure patterns, which need to be internally aligned.

3. Agree on a shared set of conventions.

The first challenge requires the agents to build an abstract layer of semantic roles and syntactic cases, as depicted in the two bottom boxes of Figure 2. Most linguists agree that those categories are not given a priori, but that they are constructed (Croft, 1991; Levin & Rappaport Hovav, 2005). A widespread assumption is that there exists some universal way for doing so (e.g. through linking rules, innate primitive structures or a hierarchy; see for example Jackendoff, 1990; Dowty, 1991; Van Valin, 2004). However, the agents in the experiments do not even get this prior knowledge. They can only recognize events based on their perceptual processes in which they detect low-level features such as which objects are moving, as described in detail by Steels & Baillie (2003). They do not dispose over any prior structures or rules that inform them about possible relations between events, for example that 'pushing' and 'pulling' both involve an actor that does something to an undergoer.

Secondly, cases do not occur in isolation but become part of larger patterns that are exploited as vehicles for expressing semantic frames such as 'X causes Y to receive Z' (Goldberg, 1995). The agents have to develop and acquire these structures as well. Moreover, cross-linguistic research has shown that case languages evolve a systematic alignment among argument structure constructions (Comrie, 2005). For instance, German typically uses the nominative case for expressing the subject in all of its argument structure constructions. The agents thus have to agree on a systematic mapping between meaning and form.

Finally, the emergent grammar needs to be propagated and sufficiently shared in the population for successful communication. The main difficulty here is that case markers are polysemous (i.e. they can take multiple meanings) and that grammar forms an abstract, intermediary layer that is not directly observable to the listener. This hurdle can nevertheless be overcome using the right kind of alignment strategy.

This paper follows the selectionist framework of language evolution introduced in the first chapter of this book (Steels, 2012) for tackling all these challenges. I assume that speakers of case languages have developed a *case strategy* that allows them to self-organize a case system. Figure 5 illustrates how that may work. The case strategy gives rise to a case system, which is exploited by the agents for producing and parsing event descriptions. The communicative outcome of a language game has a positive feedback loop to the case system and thereby influences its particular paradigmatic choices. The strategy itself is assumed to be culturally evolved as well, but the current experiments only serve to identify its components and not its origins. The feedback loop from the communicative outcome to the language strategy is thus disabled in order to ensure that the implemented strategy is adequate.

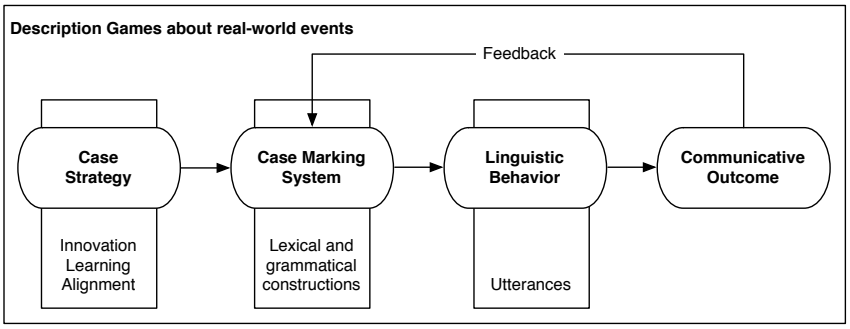

Figure 5. *Given a case strategy, a multi-agent population can self-organize a case system.*

3.1. Acquisition of a Case System

This section operationalizes the adoption and expansion components of the case strategy and demonstrates their adequacy for learning a case system through an acquisition experiment. The experiment involves a population of two agents, in which one agent is a 'tutor' who is equipped with a mature grammar system, and another agent is a 'learner' who starts with the lexical Pidgin German language but without grammar. Both agents play description games with each other in which they take turns as speaker and listener.

3.1.1. *Basic German Case*

German case is an intricate system (Müller, 2001), which intertwines three dimensions (case, number and gender) and which is notorious for its syncretism (i.e. a single case form can be mapped onto multiple, mutually exclusive values). So far, it is still not clear how learners of German manage to acquire such a complex grammatical system (Bittner, 2006; Eisenbeiss et al., 2005/2006). It is therefore a deliberate and necessary modeling choice to implement a 'basic' version of German case by excluding interferences from the gender and number systems. By doing so, the necessary learning operators for each subsystem can be studied in isolation. Once those are sufficiently understood, the full complexity of multidimensional grammatical systems can be addressed in further research.

In order to simplify the German case system, the agents only have masculine words at their disposal for describing the objects in the scenes from the puppet theater, which effectively reduces the number of case-marked definite articles used by

the agents to three: *der* (nominative), *den* (accusative) and *dem* (dative). The genitive case never occurs in the experiments. All tutor utterances are active sentences so there is a consistent mapping between semantic roles and case. The tutors have a perfect knowledge of the basic German system.

Despite the scaffold, the learning task of the learner agent is far from trivial. More specifically, he is faced with two major challenges. First, the learner has no notion of semantic roles such as 'Agent' or 'Patient' and thus has to retrieve the meaning of each case category and figure out how individual events should be organized in terms of more abstract grammatical relations. Secondly, the learner also needs to acquire the larger argument structure constructions.

3.1.2. *Adoption and Expansion*

The learning component of the case strategy provides ways to adopt a new case mapping, to extend and generalize an existing case mapping through analogy and to optimize processing by creating grammatical chunks.

Adopting case mappings. All agents are equipped with a reflective architecture for processing information at two levels (Steels & van Trijp, 2011). Suppose that the learner observes the utterance *der Junge lief* 'the boy walked'. At a first level, *routine processing* takes place in which an agent tries to process an utterance using his current linguistic inventory. At the same time, a meta-layer actively monitors the routine layer through diagnostics.

Even though the agent does not know the form *der* yet, he nevertheless parses the utterance as good as possible and comes up with a meaning that consists of three predicates: *boy(?x), walk(?ev)* and *walker(?ev, ?a)*. Even though these meanings are unconnected, the agent can infer that the boy is the walker (`?x = ?a`) from the context and is therefore able to interpret the meaning correctly.

During routine processing, the meta-layer diagnostics detected the unknown form *der*. The agent tries to make an educated guess about its function by checking whether any other problems occurred, and whether the form can be coupled to that problem. Despite arriving at a (potentially) successful interpretation of the utterance, the agent detected that he needed to make additional inferences from the context for doing so, which is reported as a problem by the following diagnostic:

- *Diagnostic 1: Detecting cognitive effort.*
  - Parse and interpret the utterance.
  - If cognitive effort is not minimal, report a problem.

Since the hearer had to make one inference in order to interpret the utterance, cognitive effort is not minimal, so the diagnostic reports a problem. As a result, one or more repairs may become active in order to try and solve the problem. Adopting a new case marker happens through the following repair:

- *Repair 1: Adopt a case mapping.* (Triggered by a problem of cognitive effort.)
  - If interpretation was successful and unambiguous and if only one inference was required:
    1. Check whether there was an unknown form in the utterance.
    2. If so, build a new case mapping and restart the comprehension task using the updated inventory.
  - If cognitive effort is too high (i.e. more than one inference) or if there are more than one uncovered forms, ignore the problem and do not adopt.

Creating a case mapping takes several steps. First, the agent builds an argument structure construction specific to the current situation. In production, this construction makes sure that a 'walker' of a walk-event is explicitly marked by assigning a *case* to it; and in parsing, the construction exploits that case information for assigning a *semantic role* to the participant that plays walker. As the agent has no semantic roles or cases yet, he builds new ones (`sem-role-1` and `case-1`). Since the argument structure construction is verb-specific, the meaning of `sem-role-1` corresponds to the walker of a walk-event, and the case category `case-1` maps onto that semantic role in a one-to-one fashion. Next, the agent makes the verbal lexical entry of *walk* compatible with the new argument structure construction by modifying its valence and stating that the participant role 'walker' can be categorized as `sem-role-1` (which may be expressed by `case-1`). Finally, the agent builds a morphological construction that maps `case-1` onto the surface form `der`.

Next, the learner tests whether the updated linguistic inventory solves the detected problems. First, the listener restarts the comprehension task and now obtains a new parsed meaning in which all the coreferential variables have been made equal: *boy(?x), walk(?ev)* and *walker(?ev, ?x)*. Next, the listener verifies whether his solution would also be adequate for production through 're-entrance' (Steels, 2003). Re-entrance is the process in which a language user 're-enters' the output of linguistic processing in his own language system. Here, the listener 're-enters' the parsed meanings in his own system in order to see simulate production so he can check how he would have verbalized that meaning himself. With

the updated inventory, the learner is now indeed capable of producing *der Junge lief*.

Extending a Case Mapping. German cases are of course not restricted to specific contexts, but they generalize across situations. The learner will therefore soon be confronted with novel contexts in which the same case marker is used. Suppose that the learner observes the utterance *der Junge ging hinein* 'the boy entered'. This time, the agent knows all forms but there is a mismatch between his hypothesis of *der* (as the marker of a 'walker') and the semantics of the event (which involves a participant entering a place). As a result, the verb-specific argument structure construction that was adopted in the previous situation fails to connect the necessary variables and the agent arrives at the following parsed meaning: *boy(?x), enter(?ev), enterer(?ev, ?a)*. The unexpressed variable equality (`?x = ?a`) is reported as a problem of cognitive effort by diagnostic 1. The agent then tries to adapt the function of *der* so that it can accommodate both uses:

- *Repair 2: Extend a case mapping.* Triggered by a problem of cognitive effort.
  - If interpretation was successful and unambiguous and if only one inference was required:
    1. Check whether a known marker is used in a novel distribution.
    2. If so, extend its case mapping and restart the comprehension task using the updated inventory.
  - If cognitive effort is too high or if there are more than one formal novelties, ignore the problem and do not adopt.

The learner assumes that the reoccurrence of a case marker in a novel context indicates that there is a semantic connection between the different uses of the marker. Since the agent has no prior knowledge about possible relations between events, he searches for *analogies* based on his perceptual experiences. Figure 6 illustrates how analogy works. The left picture shows a 'walk-to'-event, which serves as the 'source event' because the agent already knows a marker for indicating one of its roles. As can be seen, the event can be decomposed into its low-level features, which involves one moving participant and a decrease in distance between two participants until both participants 'touch' each other (i.e. their visual images overlap in perception). Analogy retains all low-level features that involve the 'source role' (i.e. the walker: the boy; indicated in black italics) and ignores the other features (gray). The algorithm then searches for corresponding low-level features in the 'target event'.

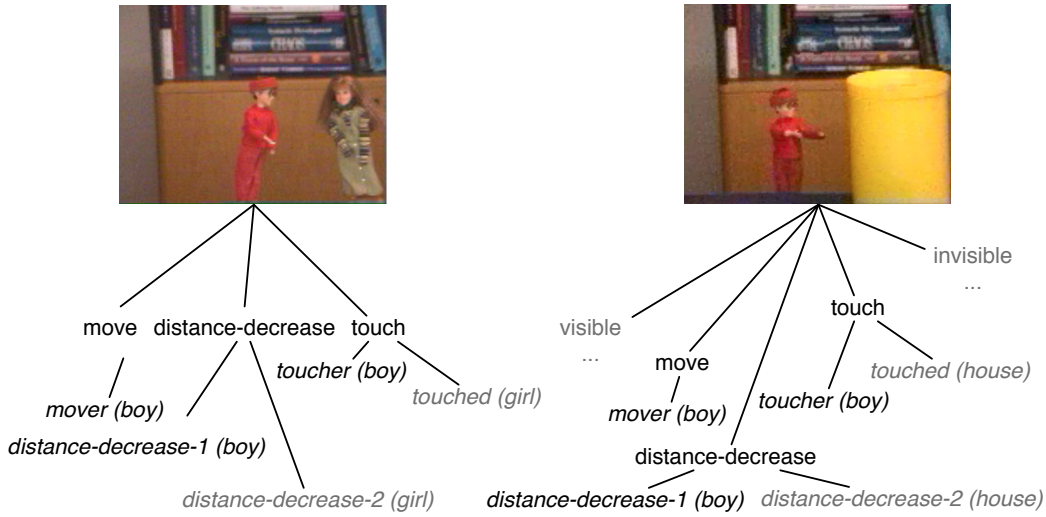

Figure 6. *Agents do not have any prior knowledge of relations between events, but they can detect analogies by mapping the event structure of a source event onto that of a target event. The left figure shows the structure of a 'walk-to'-event, the right figure shows the structure of a 'move-inside'-event. Here, the agent detects an analogy between the 'walker' on the left and the 'mover' on the right.*

The target-event ('move-inside') is shown on the right. One of the two visible participants starts moving and the distance between them decreases until both of them 'touch' each other. The event ends when one of the participants has become invisible (indicating that it has 'moved inside' the other participant). Analogy now only retains the low-level features of the target event that correspond to the features of the source role (shown in black italics) and ignores all other features. Analogy then checks whether the mapping is consistent, that is, whether the corresponding features of the target event are always filled by the same participant. Here, this is indeed the case as all corresponding features are filled by the boy and never by the house. The agent has thus retrieved an analogy between the 'walker' of the walk-event and the 'mover' of a move-inside-event. If the 'mover' was also the marked role in the utterance, the agent is now capable of semantically extending the observed case marker to accommodate for this novel usage.

If the learner fails to discover an analogical mapping, he adopts the novel use of the marker as an alternative hypothesis. During later interactions, the agent may eventually merge both hypotheses if another context provides an analogical link between the different uses of the marker. If a successful analogy is found, the learner incorporates the target role as a specific instance of `sem-role-1`. Instances are stored in memory as a cluster for later analogies and ordered according to their 'prototypicality' for the semantic role. Irrespective of whether or not the agent retrieved the analogy successfully, he modifies the semantic and syntactic combinatorial potential of the lexical constructions that were used during processing in order to make them compatible with the existing case mapping (see van Trijp, 2011a, for details). No new constructions are required.

Generalization as a Speaker. The learner can only be said to have *truly* acquired a grammar system if he is able to perform better than 'table lookup' and go beyond his training data by productively using the grammar system in novel contexts. The trigger for generalization is again a problem of cognitive effort. As a speaker, the learner agent tries to estimate whether his utterance will have the desired communicative effect on the hearer by performing re-entrance. If the speaker detects a problem of cognitive effort, he will try to solve it by extending an existing case mapping through analogy. The mechanisms for generalization are therefore exactly the same as just described above for adopting a new usage of a known case mapping.

Optimizing Processing. Even though novel case mappings may be introduced into the language system for reducing cognitive effort during *interpretation*, they also come at the risk of exploding the search space for *linguistic processing*. As explained by Steels (2011b), a language user must find a path (i.e. apply his lexical and grammatical constructions in a suitable sequence) through this search space in order to successfully produce and parse utterances. For each argument structure construction and case marker that are added to the linguistic inventory, the number of possible paths in the search space may grow exponentially, especially when competing case mappings inevitably pop up in the population.

In order to reduce the search space, the agents build larger argument structure constructions when two or three constructions are applied together during processing. The language user keeps a 'systematicity link' between the new construction and the constructions that were used for creating it. Using the network facilities of FCG (Wellens, 2011), the larger argument structure constructions have precedence

over smaller constructions during processing, which means that processing requires less steps for arriving at a solution and that there are less places in which the search can fork into different paths. The diagnostic and repair look as follows:

- *Diagnostic 2: Detecting opportunity for optimization.* If two or more argument structure constructions are applied together when processing an utterance (both production and parsing), report an opportunity for optimization.
- *Repair 2:* Triggered by an opportunity for optimization.
  - Build a new construction by combining the co-applied argument structure constructions.
  - Build a systematicity link between the argument structure constructions.

3.1.3. *Adequacy of the Adoption and Expansion Components*

Figure 7 shows the running averages of two important measures for assessing the adequacy of the case strategy: cognitive effort and learner proficiency (both on the Y-axis). The X-axis shows the number of language games on a sequential time scale. The chart shows that the learner's cognitive effort decreases to a minimum as the agent acquires the case grammar from the tutor. The other measure, learner proficiency, shows whether the learner's linguistic behavior converges with that of the tutor. The proficiency p of a single language game is 1 if the tutor agrees with the description of the learner, and 0 if the tutor disagrees. The tutor has more strict criteria for agreeing with a description and only accepts descriptions that require minimal cognitive effort for interpretation. The running average of learner proficiency rp is calculated as a Simple Moving Average over the n latest values for p (with $n = 10$):

$$\mathrm{p}(t_s) = \begin{cases} 1 & \text{if tutor agrees} \\ 0 & \text{if tutor disagrees} \end{cases} \tag{4}$$

$$\mathrm{rp}(t_s) = \frac{1}{n} \sum_{i=0}^{n-1} \mathrm{p}(t_s - i) \tag{5}$$

As can be seen in the chart, the learner's proficiency reaches 100% after 400 language games. All values in the chart show the mean values of ten series of 450 games. The error bars show standard error.

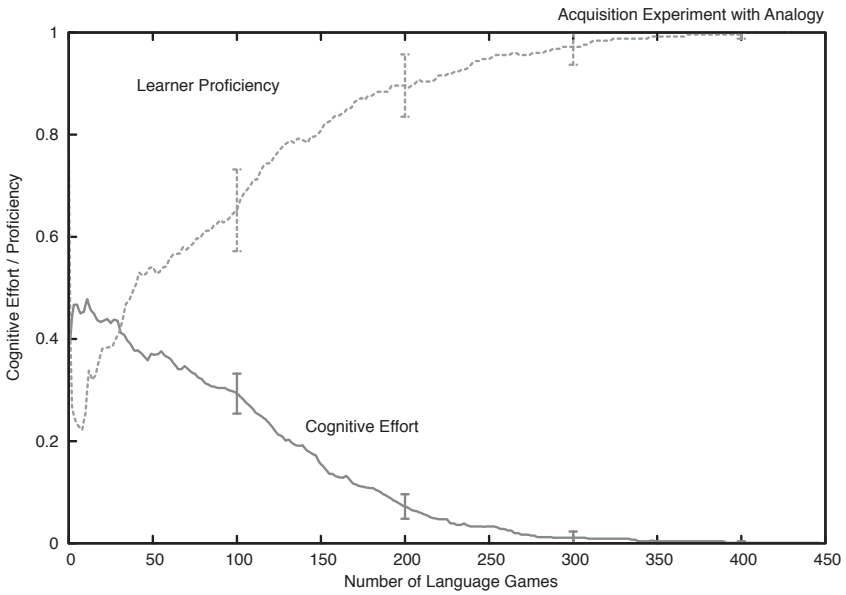

Figure 7. *The learner successfully reduces cognitive effort for interpretation as he acquires the target grammar. The learner's proficiency level reaches 100%.*

Figure 8 shows that the learner performs better than table-lookup and is indeed capable of generalizing his language system to novel situations. The X-axis shows the number of language games, and the Y-axis shows how often the learner tried to generalize his language system to novel situations when acting as a speaker. 'Correct hypotheses' are measured by counting how many times the learner successfully generalized during the last ten interactions, whereas 'wrong hypotheses' are measured by counting the number of times that the tutor did not agree with an attempt at generalization by the learner. Both numbers are divided by 10, which yields values in the [0, 1] interval. After a couple of wrong hypotheses in the very beginning of the learning experiment, the amount of correct hypotheses increases to 30%, which means that the learner correctly extended his grammar to new contexts in about three interactions out of ten, whereas the number of wrong hypotheses never surpasses 10%. In the remaining 60% of the language games, the learner did not innovate either because he already acquired the necessary constructions for describing a scene correctly, or because he considers the innovation task to be too difficult

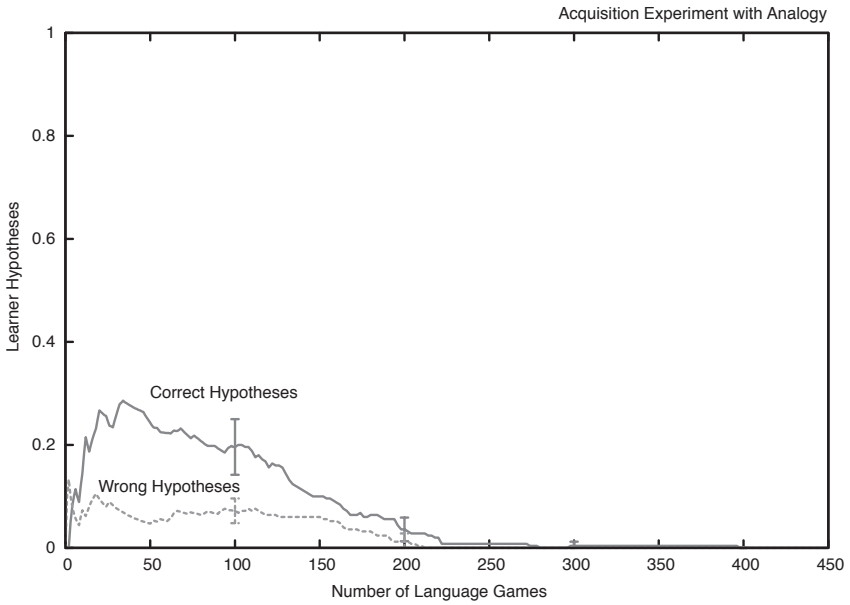

Figure 8. *This Figure shows the amount of correct and incorrect hypotheses that the learner tried out in novel contexts.*

to perform. After about 200 games, the learner only makes correct hypotheses, indicating that he has mastered the fundamentals of the German case system.

The results demonstrate that the language strategy allows the learner to acquire a language system at a fast rate: by actively taking a functional learning approach and trying out hypotheses in situated communicative contexts, the learner is often capable of one-shot learning and therefore does not need thousands of instances before he is able to generalize.

3.2. Self-Organization of a Case System

The next series of experiments feature a population of 10 agents that start with a Pidgin language (consisting of English words). Without any tutors in the population, the agents have to self-organize their case system.

3.2.1. *Innovation and Alignment*

The case strategy uses the same adoption and expansion components as described in section 3.1.2 for both learning and innovation. As opposed to most other models on case (see for example Moy, 2006), innovation is assumed to be *speaker-based*. If the speaker fears that the hearer might face a problem of cognitive effort (which he can detect through diagnostic 1 when re-entering his utterance), he tries to solve it through repairs 1 and 2. The only difference is that instead of *adopting* a new case mapping (repair 1), he may invent a new one. Inventing a new case mapping is, however, risky because the speaker needs to assume that the hearer retrieves the function of the new case marker. Repair 2 is therefore always preferred: by recruiting an analogous case in a novel context, the hypothesis space of the hearer is reduced, which means that there is a higher chance that the hearer is able to adopt the innovation. The repair also requires less changes to the linguistic inventory and therefore less memory.

The alignment component allows the agents to handle the inevitable variation that pops up in multi-agent populations and agree on a shared set of conventions. Here, alignment is implemented using lateral inhibition dynamics (as first introduced by Steels, 1995), which means that every element of the linguistic inventory may have a 'dynamic confidence score' that reflects how confident the agent is that the element is indeed conventional in his speech community. After each linguistic interaction, agents update these preference scores based on the communicative outcome of the language game: they increase the score of items that are eligible for reward (for instance a successful case marker), and they decrease the score of all the competitors of those items (e.g. a rivalling case marker).

There are two main problems that need to be addressed when dealing with grammatical alignment. First, grammatical categories are polysemous and multifunctional because they generalize across meanings. Instead of having direct, one-to-one competitors (as typically found in for instance Naming Game dynamics; Steels, 1995), they compete with other categories for settling the boundaries between them (see Figure 9). Alignment therefore needs to more fine-grained than the lateral inhibition dynamics typically found in lexical alignment (ibid.). The second problem is that grammatical categories do not occur in isolation, but as part of a grammatical system comprising multiple constructions. In order to ensure 'systematicity' (i.e. the systematic reuse of categories across constructions), alignment should therefore not treat entries of the linguistic inventory as individual items, but favor *families of constructions*.

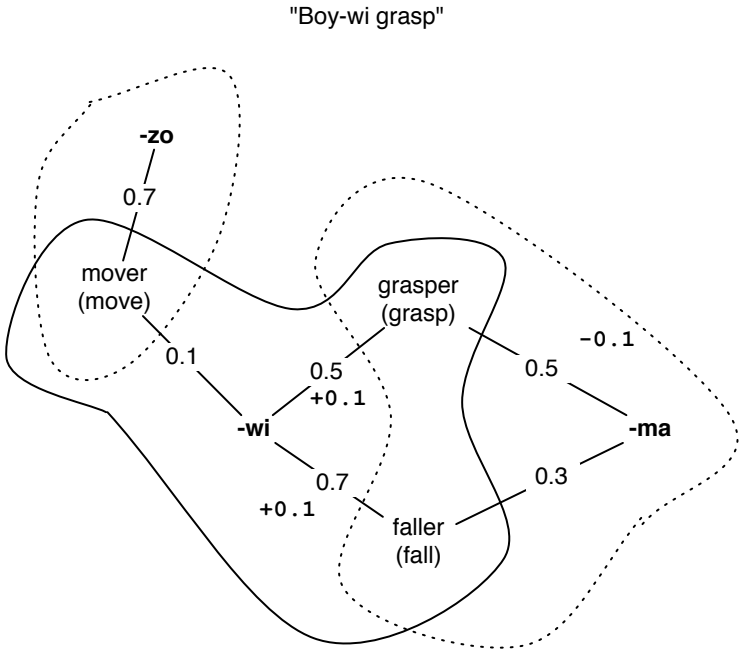

Figure 9. *Polysemous categories do not engage in a one-to-one competition but rather have to fight over their boundaries. This example shows how the marker* -wi *competes with* -zo *and* -ma *for covering various participant roles.*

The first challenge is addressed by *shaping* the meaning or function of a case category (Wellens et al., 2008). Figure 9 illustrates how shaping can be achieved. The Figure shows three case markers: *-wi*, *-ma* and *-zo*. The markers themselves are not scored, but rather their *links* to the participant roles that they can express. For example, *-wi* can be used for marking a faller, grasper or mover in three events (hence *-wi* has a type frequency of three). Each link has a confidence score between 0 and 1, which reflects how confident an agent is that the mapping between the marker and the participant role is a convention in his speech population. Suppose that the hearer successfully observes the utterance *Boy-wi grasp* 'the boy grasped (something)'. The agent first of all rewards the use of *-wi* as a marker for expressing the grasper of a grasp-event by increasing the confidence score by 0.1. The agent also performs lateral inhibition and punishes all competing links: the score of the link between the marker *-ma* and the grasper-role is therefore decreased by

0.1. So far, alignment works the same as in lexical evolution. However, on top of the standard lateral inhibition dynamics, the agent also rewards all other links in the network of *-wi* for which it is the dominant one: the score of the link the faller-role and *-wi* is increased by 0.1. The score of the link to the mover-role, however, remains unchanged because there is a more dominant competitor (*-zo*). This alignment strategy favors more polysemous grammatical categories while at the same time allowing fine-grained competition; and it 'shapes' the meaning of a case category over time.

The second challenge addresses the question of how a case category can become dominant *across* constructions in a structured linguistic inventory. This challenge is solved through *multi-level alignment* (Steels et al., 2007; Steels & van Trijp, to appear), which is here applied to the confidence scores of constructions themselves. In multi-level alignment, not only successfully applied constructions are eligible for reward, but all of their systematically related constructions as well. In the current example, a monovalent construction was used containing the marker *-wi*. Multi-level alignment now implies that all other constructions that use the same case mapping (for example bivalent constructions in which the case mapping is part of a larger pattern) are rewarded as well. In each case, the direct competitors are punished through lateral inhibition. Multi-level alignment thus favors *families* of constructions over individual categories, and case mappings therefore have a higher survival chance if they become part of a successful family.

In sum, the above alignment strategy favors cases that are more general (or have a higher type frequency) and which form part of successful families of constructions. There is plenty of evidence from natural languages that this alignment strategy indeed plays a role in the evolution of case. For example, Barðdal (2011) suggests that accusative subject predicates in Icelandic are attracted by the dative subject construction because it has similar semantics but a higher type frequency. Likewise, Barðdal (2009) shows that nominative is the most productive case in Germanic languages because it has the highest type frequency and the highest semantic distribution of all cases.

3.2.2. *Adequacy of the Case Strategy*

This section first describes some system properties of the emergent languages and then compares the results to case systems in natural languages.

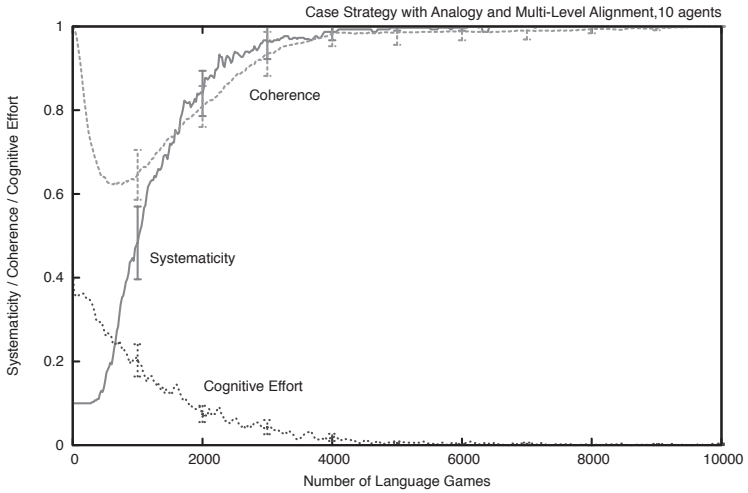

Figure 10. *This graph shows that the agents succeed in reducing their cognitive effort, while converging on the same conventions and maintaining a fully systematic case system.*

Cognitive Effort, Coherence and Systematicity

Figure 10 shows the running average of cognitive effort and the evolution of coherence and systematicity in populations of 10 agents, equipped with analogy and multi-level alignment. The values average over ten series of 10.000 games on a sequential time scale; error bars indicate standard error. The chart shows that the agents succeed in reducing the cognitive effort needed for semantic interpretation to a minimum after 5.000 games (or 1.000 on average per agent).

The *coherence* measure tracks the linguistic behavior of the agents and indicates whether they all use the same forms for expressing a particular 'participant frame'. A participant frame consists of a number of participant roles, such as {walker} (for utterances equivalent to *someone walks*) or {walker, destination} (for utterances equivalent to *someone walks to some place*). The coherence measure keeps track of the last ten forms used for expressing each participant frame. For each frame, coherence is calculated by taking the most frequently used form and dividing the number of times that form was used by 10. Population-wide coherence averages over all the individual coherences. As can be seen in the chart, the agents behave

in a coherent way in the beginning of the experiments because they start out with a shared Pidgin language. As the agents start introducing case markers, however, the variation in how meaning is mapped to form varies in the population, and coherence drops to 60%. The multi-level alignment strategy allows them to quickly restore coherence almost to its maximum value after 4.000 language games (800 on average per agent). After 9.000 games, all agents agree on the same linguistic behavior.

The third measure, systematicity, shows whether the same meanings are systematically expressed using the same forms *across* constructions, for example whether the 'giver' is marked by the same case regardless of which participant frame it occurs in (i.e. regardless of whether it occurs in a monovalent, bivalent or trivalent construction). Here again, the alignment component of the case strategy guarantees a maximum result: the agents start with low systematicity, but gradually align to a systematic distribution of cases across different argument structure constructions and end up with a fully systematic language.

Systematicity is measured by taking the most frequently used form for expressing a particular participant frame, and checking whether the same forms occur in all of the forms used for expressing participant frames that overlap in meaning. If so, a score of 1 is counted. If not, a score of 0 is counted. The sum of all these scores is divided by the number of participant frames in the meaning space, which yields a score between 0 (no systematicity) and 1 (maximum systematicity). For example, suppose that the participant frame {appearer} is most frequently marked by *-bo*, {appeared} by *-ka* and the combination of the two as *-bo -si*. First we take {appearer} and check whether its marker also occurs in the combination with 'appeared': this is indeed the case so the form-meaning mapping is systematic in both constructions, which is counted as '1'. For {appeared}, however, the pattern uses a different marker *-si* so no systematic relation exists across constructions, which is counted as 0. The combination itself does not occur in a larger pattern so it is not considered by the systematicity measure. In this example, systematicity would thus be 0.5.

Expressivity and Learnability

The second hypothesis that needs to be investigated is how grammar can lead to greater expressivity and learnability. An important indicator of expressivity is the size of the linguistic inventory. The smaller the size for a given meaning space, the more expressive a language system is. Figure 11 compares the expressivity in two different set-ups: the top chart shows a population capable of inventing new

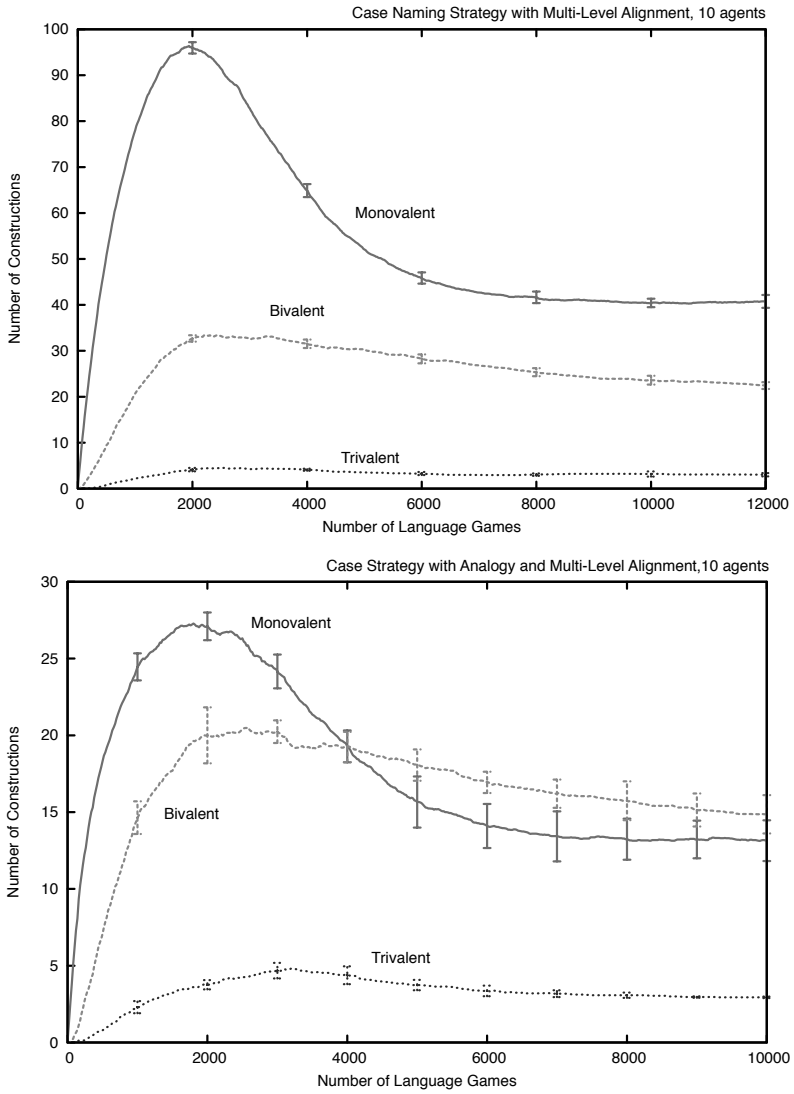

Figure 11. *The top graph shows the inventory size if the agents do not exploit analogy and therefore invent markers for every possible participant role. The bottom graph demonstrates that by reusing existing markers through analogy, the size of the linguistic inventory is minimized. The graphs average over ten series of language game experiments in each population; error bars indicate standard error.*

cases but not capable of generalization through analogy. The bottom chart shows the results of a population equipped with analogical reasoning. Both charts show the average inventory size per construction type. The average inventory size I for each time step t_s is calculated by dividing the sum of the length of the language system of each agent $\mathscr{L}_{a(t_s)}$ by the population size N:

$$\mathrm{I}(\mathrm{t}_s) = \frac{1}{N} \sum_{a=1}^{N} |\mathscr{L}_{a(t_s)}| \tag{6}$$

The top graph shows that, without generalization, the case strategy almost corresponds to a Naming Strategy whereby agents can invent 'names' for marking each participant role without reusing them through analogy. The agents innovate new case forms during the first 2.000 interactions, leading to an overshoot in the population of almost 100 'case names' for 40 participant roles. This overshoot is followed by an alignment phase in which the population settles on 40 case names. The overshoot for bivalent and trivalent constructions (which combine case markers used in monovalent constructions) is less pronounced, because they are typically introduced when some case markers have already become dominant in the population, hence there is less variation in innovation. The bottom graph shows that analogy drastically reduces the required inventory size: the number of monovalent constructions never exceeds 30 at the same time, and different populations settle on 14 markers on average. The effect on bivalent and trivalent constructions seems to be negligible.

Figure 12 represents the competition among case markers within a single agent in one of the experimental runs. The Y-axis shows the type frequency of each construction, which is simply the number of participant roles that a case marker can potentially express. The x-axis shows the number of language games on a sequential time scale, so each agent plays on average 2.000 games. As can be seen, the agent has to track 36 competing case markers and adapt his inventory to changes in the population. Most of the markers die halfway the experimental run. In the end, the population settles on three main case markers (with respective type frequencies of 8, 9 and 10 participant roles) that almost take up three quarters of the meaning space.

The Emergent Language Systems

It is worthwhile to step into the shoes of a linguistic anthropologist and take a deeper look at the case systems that evolve in the populations. Figure 13 shows the

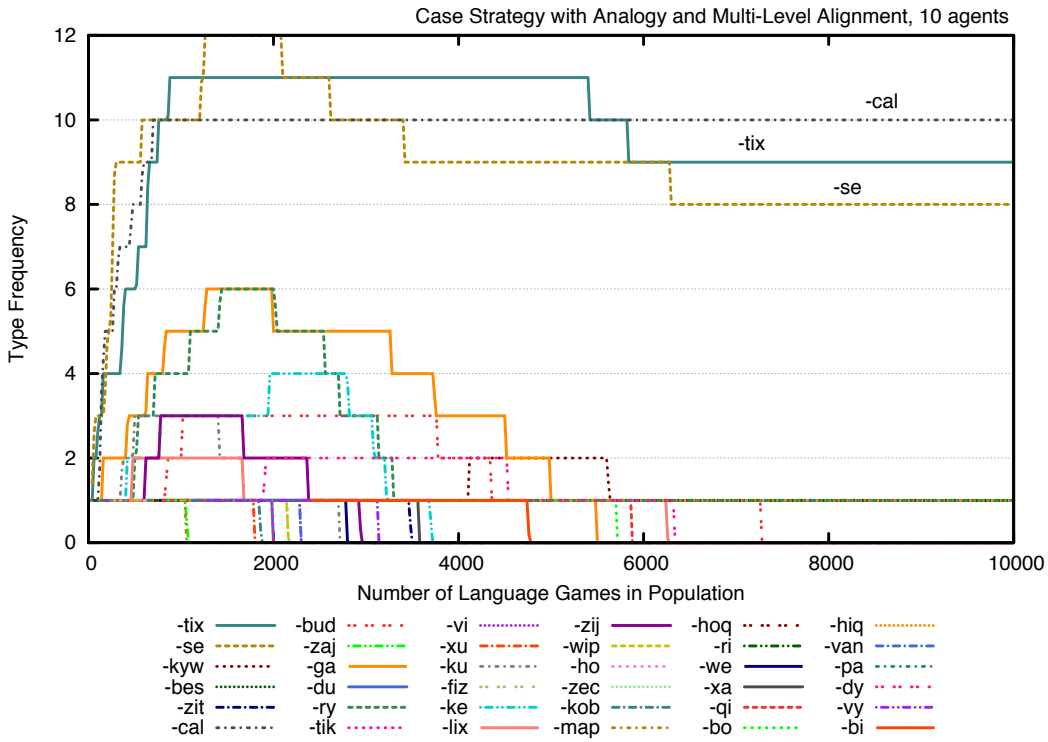

Figure 12. *This Figure looks inside the head of one agent during one experimental run. It shows how the agent keeps track of competing case markers over time.*

case mappings that have been developed in one population. As can be seen, the population has settled on four, semantically quite coherent cases and seven participant-role-specific cases. The markers *-tux* and *-hiq* are used for agent-like roles, whereas *-pav* and *-my* take on patient-like and locative-like roles. Interesting is that the difference between *-pav* and *-my* often involves direction: whereas *-pav* often involves a destination or proximity, *-my* marks role that typically involve outward movement. Here are some example utterances to illustrate the language:

(7) *Boy -tux move-inside house -pav*
boy case-12 entered house case-5

'The boy entered the house.'

(8) *Jack -my go-away Jill -tux*
Jack case-10 walk away from Jill case-12

'Jill walked away from the boy.'

(9) *Jack -tux grasp blue -pav*
Jack case-12 grasped the blue block case-5

'Jack grasped the blue block.'

(10) *girl -hiq separate blue -my*
the girl case-7 let go the blue block case-10

'The girl let go off the blue block.'

The effect of analogy and the relevance of the results for linguistics becomes more apparent when the emergent languages of different experimental runs are compared to each other. One current discussion, for example, is the status of 'semantic maps' (also known as 'conceptual space') in linguistic typology. Some researchers hypothesize that semantic maps represent a universal and contiguous conceptual space that underlies all languages, so languages only differ in how they carve up that space (Haspelmath, 2003). Other researchers argue that semantic maps offer a valuable research methodology without subscribing to any hypotheses about their universality (Cysouw, 2007; also see van Trijp, 2010 (for more on this discussion).

The results presented in this paper suggest that semantic maps are not necessarily universal, but that they can emerge as a side-effect of exploiting analogy for semantic extension. Figure 14 shows a primitive semantic map that compares two case categories that have been developed independently by two different populations. The left case category (marked by *-pav*) shows a significant overlap with *-cat*. Since *-pav* is taken from the language system depicted in Figure 13, it is possible to see how *-cat* relates to those other categories as well. It turns out that *-cat* is a more general category than *-pav*. Interestingly, it does not make the same split between direction of motion as opposed to the case system in Figure 13.

When comparing all the emergent language systems to each other, it may look from the outside as if the agents were given a contiguous semantic map in advance. However, all they had at their disposal was the input from visual processing extracted from actions in their world. In other words, the experiments demonstrate that semantic maps can be dynamically formed as the result of the distributed processes whereby the agents shape and continue to reshape their language.

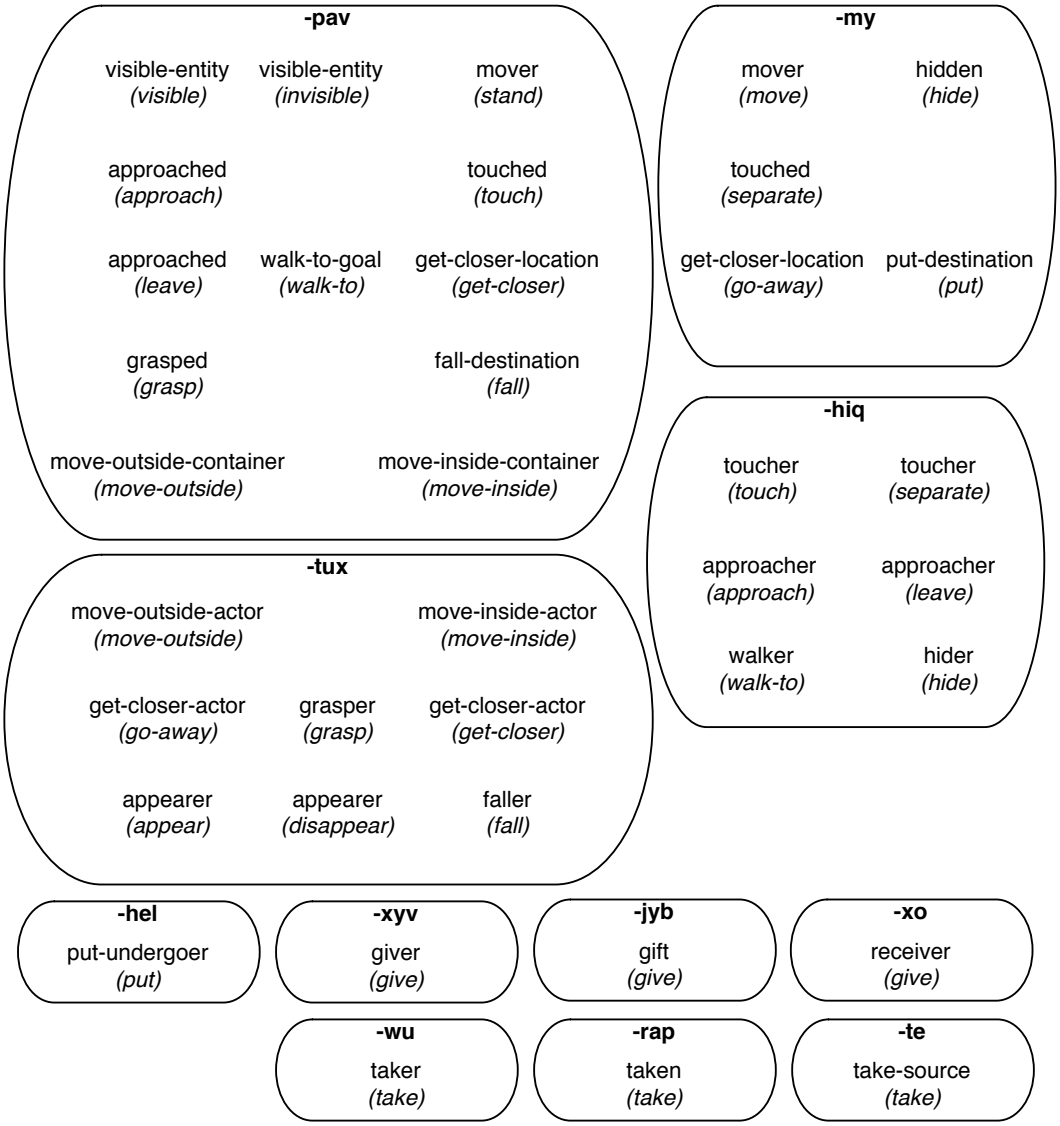

Figure 13. *This Figure shows the emergent case system of one of the populations. For each case marker (bold-faced), it shows which participant roles they can be associated with. The events in which those participants play a role are shown between brackets.*

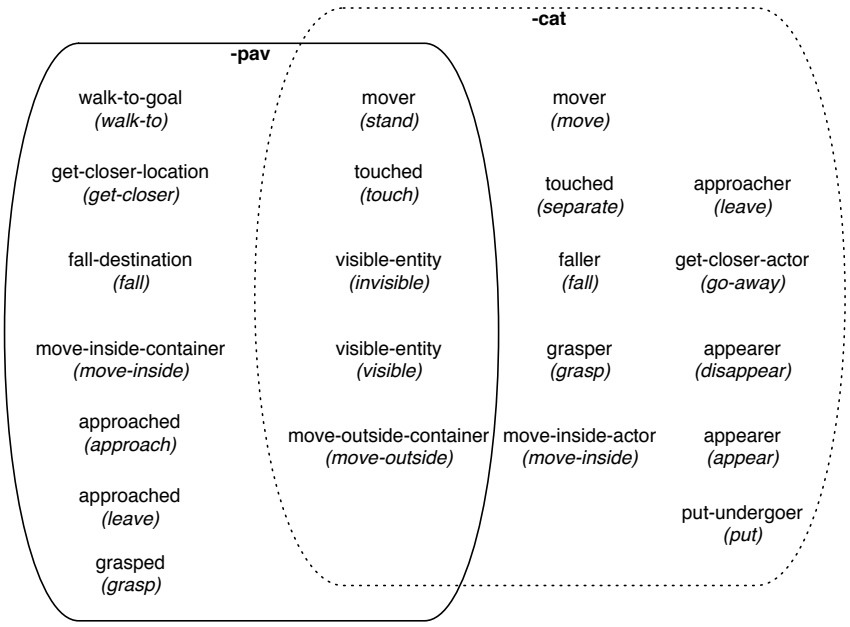

Figure 14. *Semantic maps emerge as a side-effect of exploiting analogy for semantic extension.*

4. Why and how do case systems evolve new functions?

Any theory of language evolution does not only need to explain why and how a language system may emerge, but also why such systems may decay over time and undergo drastic changes. This section focuses on a widely observed phenomenon in natural languages, namely that a fossilized language system gets recruited for different purposes. The hypothesis explored here is that a language system may develop novel functions if its original selective advantages for communication have been taken over by rivaling systems.

4.1. Reconstruction of Spanish Pronoun Systems

In order to make the discussion concrete, the experiments reported in this section draw inspiration from an ongoing evolution in the Spanish personal pronoun system, which forms a striking puzzle for many theories on language change. Spanish is a romance language that, like many other languages descending from Latin, has lost

its case system. The Spanish personal pronoun system (see Table 1) still shows traces of a previous case declension, which is most apparent in the third person that differentiates among nominative, accusative and dative. The third person also features a gender distinction between masculine and feminine in the nominative and accusative cases, but not in the dative case. This pronoun system is known by scholars as the 'etymological system' (Fernández-Ordóñez, 1999).

	Nominative	Accusative	Dative
1st Person	*yo*	*me*	*me*
2nd Person	*tú*	*te*	*te*
3d Person Masc.	*él*	*lo*	*le*
3d Person Fem.	*ella*	*la*	*le*

Table 1. *Singular pronouns in Standard Spanish.*

The etymological system is gradually changing into a 'referential system' in which the accusative and dative cases are collapsed. The system, however, is not impoverished but starts to differentiate gender, number and noun class. 'Referential system' is in fact an umbrella term that covers many variations of the system, which are also not uniformly propagated in the regions of Spain and which are still in competition with the etymological one. Depending on the particular flavor of the referential system, the variations are called *leísmo*, *laísmo* and loísmo (Fernández-Ordóñez, 1999).

Leísmo refers to the use of the pronoun *le* (etymologically a dative pronoun) instead of accusative *lo*, most frequently for marking singular, masculine persons (ex. 11). This use is also the only one that is accepted by the Real Academia Española (RAE).The *Laísmo* variation denotes the use of *la* (etymologically an accusative pronoun) instead of dative *le* with a feminine referent (ex. 12). Finally, *Loísmo* is the use of *lo* (etymologically accusative) instead of dative *le* with masculine or neuter referents (ex. 13):

(11) *Le vi (a Javier).*
her.ACC saw-1P.SG.PAST (Person-Marker Javier)

'I saw him (Javier).'

(12) *La dio un regalo (a Maria).*
her.DAT gave-3P.SG.PAST INDEF.ART.M present to Maria

'He gave a present to her (Maria).'

(13) *Lo dio un regalo (a Juan).*
him.DAT gave-3P.SG.PAST INDEF.ART.M present to John

'He gave a present to him (John).'

The mystery of an evolution as the Spanish one is the following: Spanish people are *aware* of the existing variation in their language. They can perfectly understand speakers from different dialect groups and often make fun of each other's different paradigmatic choices. This observation bluntly contradicts the widespread assumption that linguistic change occurs because of misunderstandings or mismatches in learning when language is transmitted from one generation to the next (see Haspelmath, 1998; Croft, 2000, for similar arguments). So why is it that Spanish speakers, who are still taught the etymological system in school, prefer to use their pronouns for a different use? And how do they manage to do so?

4.1.1. *Reconstruction Experiment*

The questions raised in the previous section can again be answered by taking an evolutionary point of view. I hypothesize that the remnants of the case system in Spanish pronouns are exploited for different functions because their original selective advantage for communication has been usurped by other systems: the cognitive load of finding out who is doing what to whom is now sufficiently carried by word order, subject-verb agreement and pronouns that the existing case distinctions have lost their primary functionality.

The hypothesis can be verified through a reconstruction experiment in which the current variation in Spanish is modeled. Variation is a tough challenge for computational formalizations, because traditional approaches typically implement an idealized, homogeneous state of a language in which variation is considered to be ungrammatical. A successful implementation has nevertheless been achieved in Fluid Construction Grammar in which speakers can both parse and produce all competing systems, yet having their own individual preferences.

The implementation follows the same approach as described in section 2.3 for German, and uses a Fluid Construction Grammar operationalization as proposed by van Trijp (2011a,b). For the current purposes, two syntactic features are of importance: case and gender. The diagram in Figure 15 highlights these features for the etymological system (left) and the referential systems (right). Each box in the diagrams shows information for a particular pronoun and should be read as follows: the linguistic representation of the pronoun (for example *lo*) has a feature called `syn-cat`, whose value 'includes two unique features' (`==1`) called `case` and `gender`. The feature `gender` takes a symbol as its value: `M`(asculine), `F`(eminine)

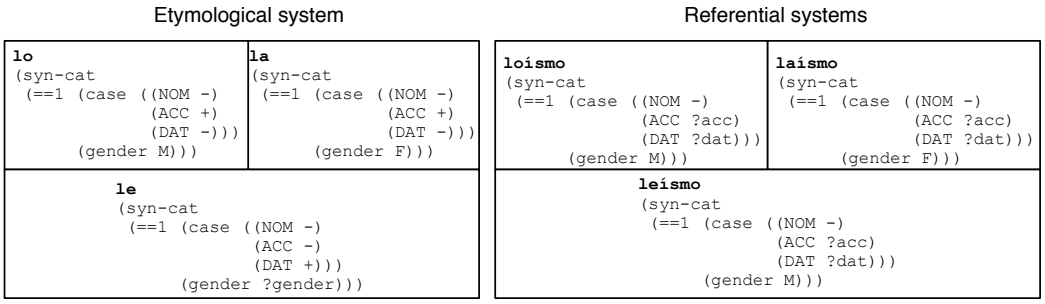

Figure 15. *Left: the case and gender features of three pronouns in the etymological system. Right: three variations of them as they occur in the referential systems.*

or a variable `?gender` (indicated by a question mark) if gender is unspecified (as in the etymological *le*). The value of the feature `case` is a 'feature matrix' (van Trijp, 2011b), which is an array of feature-value pairs that represent a case paradigm. In the etymological system, the feature matrices for *lo* and *la* indicate that they are accusative pronouns (indicated by '+') that exclude nominative or dative uses (indicated by '–'), whereas *le* is always dative. The referential systems on the right show small changes with respect to the etymological system. Both the *loísmo* and *laísmo* variation keep their gender value but are now underspecified for accusative and dative. The *leísmo* variation is also underspecified for those two cases, but specified as a masculine pronoun.

In order to model the variation, agents are equipped with the etymological system and with the referential systems, which allows them to parse and produce all variations. Each pronoun is given a preference score, so individual speakers prefer one system over the other. Figure 16 provides a snapshot taken from the FCG implementation, which illustrates a speaker who wants to say 'He/she gave a present to him' in Spanish. The Figure summarizes a 'production task' (`task-56`) by the agent, which consists of various processes such as conceptualizing a meaning to express, starting production (`create-initial-structure`), applying constructions (`produce-lex`, `produce-con` and `produce-morph`) and rendering the result of production into an utterance. As can be seen, the production task splits into two hypotheses (`task-56-1` and `task-56-2`) during the process `produce-morph`, in which the agent applies his morphosyntactic constructions.

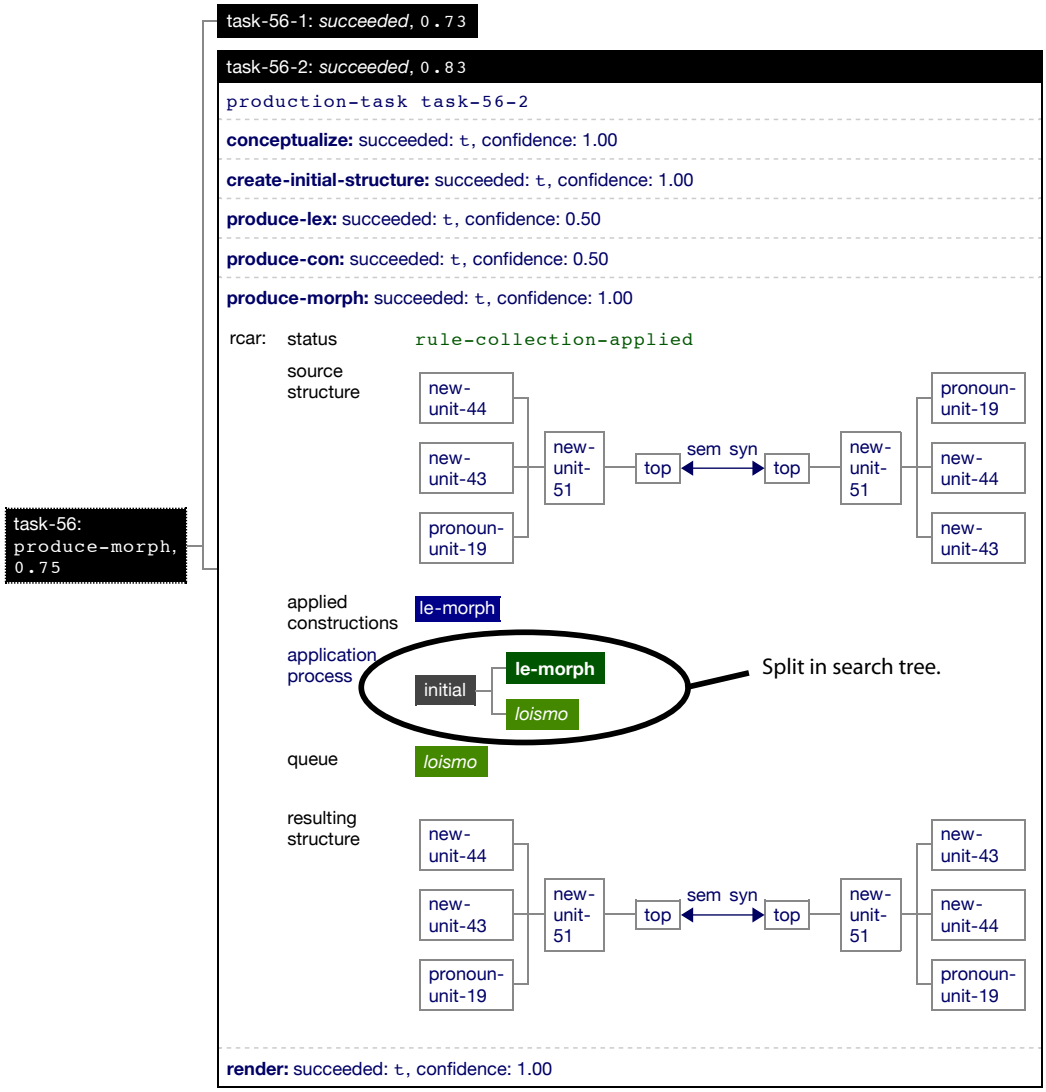

Figure 16. *The FCG implementation is capable of coping with variation in the pronoun systems in Spanish. Here, the speaker prefers the etymological system over the alternative loísmo for producing* Le dio un regalo *'He/she gave a present to him'. The speaker is, however, aware of the existing variation as it causes a split in search during linguistic processing.*

When inspecting that process for `task-56-2` (the opened box on the right in the Figure), we can see what caused this split. It is not necessary to understand the details of the Figure (interested readers can check Steels, 2011a, for more on FCG); for the current purposes, only the application process of morphological constructions matters. As can be seen, the speaker has a choice between verbalizing the recipient as *le* (standard Spanish) or *lo* (loìsmo) in the application process. Here, the speaker has a higher preference score for *le* and hence produces the utterance *Le dio un regalo*. At the same time, the speaker is aware of the existing variation because the split in the search tree can be detected through the meta-layer of diagnostics and repairs.

4.2. Demonstration of a Paradigm Shift

The experiments reported in this section implement a scenario of a case system in peril. A population of 10 agents is equipped with the etymological system of Spanish, whose case system has become fossilized. As before, the agents play description games with each other.

4.2.1. *Language Strategies*

The agents have the same case strategy as illustrated in the previous experiments, with the difference that its innovation component is no longer productive. This means that the agents can still acquire case distinctions, but that they will no longer invent new ones because the cognitive load of retrieving the event structure of an utterance is now carried by word order constructions. At the same time, the agents do have an active Gender strategy, as all new nouns in Spanish are gender-marked. Gender has also been proposed by linguists as a possible explanation for the current changes in Spanish (Fernández-Ordóñez, 1999). The strategy has the following diagnostics and repairs:

1. *The speaker wants to express gender*
    - Diagnostic: Gender was not expressed explicitly.
    - Repair: The speaker recruits a gender-marked pronoun (e.g. *lo* or *la*) or he reanalyzes the underspecified pronoun (*le*) to become gender-marked.
2. *The hearer encounters innovative use*

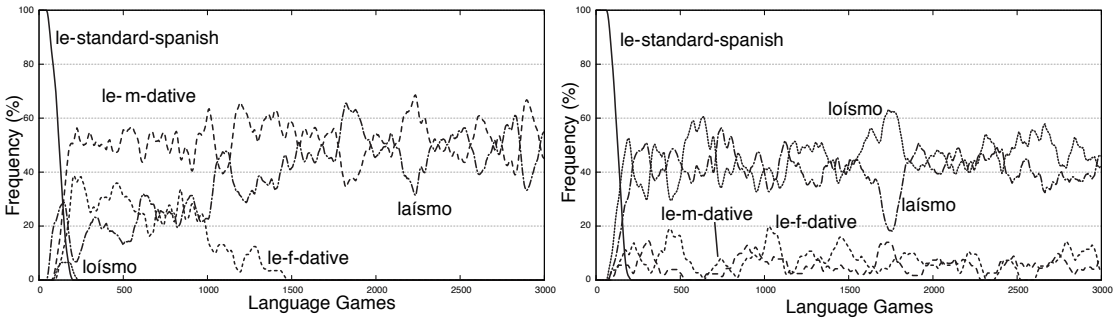

Figure 17. *All populations arrive at a paradigm shift in their pronoun system.*

- Diagnostic: The hearer detects a mismatch between the function of a pronoun and the function that was expected by an argument structure construction.
- Repair: The hearer adopts the new usage of the pronoun.

3. *The speaker encounters additional search*

- Diagnostic: The speaker detects splits in the search during linguistic processing.
- Repair: The speaker introduces additional gender constraints to avoid this search in the future.

The gender strategy's alignment component increases the confidence score of a pronoun if it was used in a successful interaction, and punishes all alternative pronouns that could have applied as well.

4.2.2. *Experimental Results and Discussion*

Ten series of language games experiments were conducted. Given a fossilized case strategy and a dominant gender strategy, all populations were capable in shifting their case-based pronoun system to a referential system without loss in communicative success. Each time, a referential system is selected because it increases grammatical coherence and therefore the learnability of the language.

The two graphs in Figure 17 show two different experimental outcomes. The x-axis shows the number of language games played in the population and the Y-axis

shows the frequency of use of a pronoun. The left graph shows a population that has evolved the *laísmo* variation. In the beginning of the experiment, both *la* and *lo* start to compete with *le* in order to mark gender in the dative case. *Le* is only able to survive this competition because it is reanalyzed as a masculine pronoun, but it has to concede all feminine referents to *la*. In this population, the three-way case distinction is thus maintained for masculine uses only. The right graph shows a population that has completely collapsed the accusative-dative distinction, as it evolved a referential system that combines *loísmo* and *laísmo*.

In none of the experiments does the *leísmo* variation occur even though in real life it is the most frequent one (and the only variation that is officially accepted by the RAE). The reason is that there is already a gender distinction in the accusative case, hence the gender strategy does not pressure the agents to make changes here. Indeed, most occurrences of *leísmo* in Spanish actually employ *le* as a pronoun for personal, animate and volitional referents, which are semantic features, rather than as a marker for syntactic gender. Empirical data also shows that *lo* is infrequent in Spanish for marking syntactic gender, and mainly occurs for marking a distinction between mass and count nouns (Fernández-Ordóñez, 1999). In order to provide a more accurate model of the ongoing evolution in Spanish, it is therefore necessary to implement a Strategy for Nominal Reference rather than a Gender Strategy. This observation, however, does not invalidate the results obtained in this experiment, which demonstrates how a grammatical system can get recruited by a new language strategy if its original functions have been taken over by competing systems.

The fact that different experimental runs yield different language systems shows how the same language strategy and communicative pressures can still lead to different paradigmatic choices. Furthermore, we see an effect of 'canalization' in the sense that the possible evolutionary pathways are restricted by system choices that have already been made in the language. This explains why different populations often ended up making the same paradigmatic choices despite variation in the population and lack of contact between different populations.

5. Discussion and Conclusions

Case has long been considered as being too complex to be either learned or to be functional for communication. This paper, however, has shown that many mysteries concerning case can be revealed by taking the viewpoint of evolutionary linguistics. Even though further research efforts are required, this paper has made several contributions to the study of case, whose validity has been demonstrated

through language game experiments and whose hypotheses can be further examined in other research disciplines.

First, I have shown *why* case may emerge in a language: it offers a selective advantage for communication because it reduces the cognitive effort needed for interpretation while at the same time limiting the required resources for doing so. This claim can be empirically validated through psycho- and neurolinguistic experiments. First work in these disciplines indeed seems to confirm the positive effect of case on processing (Clahsen et al., 2001). The same hypothesis has also yielded interesting results using mathematical models (Jäger, 2007). Future research in evolutionary linguistics needs to investigate how the selective advantages of case systems differ from those of other strategies for marking event structure such as word order or verbal agreement, and how the choice for one strategy may impact the possibilities of other choices in the language.

The second part of this paper demonstrated *how* a population can self-organize a case system through cultural language evolution and has produced the most promising results in the domain of case so far. The experiments demonstrated that language users need far less prior knowledge than is commonly assumed in linguistics for constructing abstract categories such as semantic roles. By exploiting general cognitive functions such as analogical reasoning or similarity-based categorization, such grammatical categories can be directly grounded in perception. In the current experiments, perception has been limited to visual experiences, but advances in robotics may lead to experiments that involve sensorimotor experiences as well. The results also indicated that supposed language universals (such as semantic maps) can emerge as a side-effect of locally distributed communicative interactions. Of course, the experiments also raised new questions that yet have to be explored. The first question is how agents can develop a language strategy for case from scratch. Another unsolved problem is why case systems in natural languages are more abstract (and semantically less coherent) than the artificial languages reported in this paper.

Finally, the same methodology has advanced our understanding of why and how language systems may become recruited for different functions and how existing paradigmatic choices influence the new choices made in a language. Through a case study on Spanish, it was shown how variation, which causes fundamental problems for some theories of language change, can be explained through an evolutionary account. Future work in this area needs to couple the experiments on the emergence of case and the experiments on strategy competition. In order to achieve this integration, however, it is necessary to gain a deeper understanding in how and why

case systems may decay and eventually disappear in a population, as documented by empirical linguistic studies (Barðdal & Kulikov, 2009).

Acknowledgements

The research reported in this paper was conducted at and funded by the SONY Computer Science Laboratory in Paris and the VUB AI-Lab at the Vrije Universiteit Brussel. Additional funding came from the FP6 EU project ECAgents and the FP7 EU project ALEAR. I wish to thank Luc Steels, director of both labs in Paris and Brussels, for his constructive feedback and my colleagues for their support (in particular Joris Bleys for the many fruitful discussions). I thank David Gil (Max Planck Institute Evolutionary Anthropology Leipzig, Germany) for providing examples from Riau Indonesian. I also thank Johanna Barðdal (University of Bergen, Norway), Tao Gong (Max Planck Institute for Evolutionary Anthropology Leipzig, Germany) and Paul Vogt (Tilburg University, The Netherlands) for reviewing this paper and thereby significantly improving its quality. Special thanks also go to Mar Garachana and Joe Hilferty (University of Barcelona) and Javier Valenzuela (University of Murcia) for providing me with the necessary insights into the Spanish data and for the great collaboration. I also thank my other informants for German and Spanish: Katrien Beuls, Vanessa Micelli and Michael Spranger. All remaining errors are of course my own.

References

Amberber, Mengistu, Helen de Hoop (Eds.) (2005). *Competition and Variation in Natural Languages: The Case for Case*. Oxford: Oxford University Press.

Barðdal, Johanna (2009). The development of case in Germanic. In Johanna Barðdal, Shobhana Chelliah (Eds.), *The Role of Semantic, Pragmatic and Discourse Factors in the Development of Case*, 123–159. Amsterdam: John Benjamins.

Barðdal, Johanna (2011). Rise of dative substitution in the history of Icelandic: A diachronic construction grammar approach. *Lingua*, 121, 60–79.

Barðdal, Johanna, Shobhana Chelliah (Eds.) (2009). *The Role of Semantic, Pragmatic and Discourse Factors in the Development of Case*. Amsterdam: John Benjamins.

Barðdal, Johanna, Leonid Kulikov (2009). Case in decline. In Andrej Malchukov, Andrew Spencer (Eds.), *The Oxford Handbook of Case*, 470–478. Oxford: Oxford University Press.

Bittner, Dagmar (2006). Case before gender in the acquisition of German. *Folia Linguistica*, 40(1–2), 115–134.

Blake, Barry (2001). *Case*. Cambridge Textbook in Linguistics. Cambridge: Cambridge University Press, 2nd edn.

Butt, Miriam (2006). *Theories of Case*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Carstairs-McCarthy, Andrew (2004). What morphology can tell us about grammar. In Miriam Butt, Tracy Holloway King (Eds.), *Proceedings of the LFG'04 Conference*. Stanford: CSLI Publications.

Clahsen, Harald, Meike Hadler, Sonja Eisenbeiss, Ingrid Sonnenstuhl-Henning (2001). Morphological paradigms in language processing and language disorders. *Transactions of the Philological Society*, 99(2), 247–277.

Comrie, Bernard (2005). Alignment of case marking. In Martin Haspelmath, Matthew S. Dryer, David Gil, Bernard Comrie (Eds.), *The World Atlas of Language Structures*, chap. 98–99. Oxford: Oxford University Press.

Croft, William (1991). *Syntactic Categories and Grammatical Relations. The Cognitive Organization of Information*. Chicago: Chicago UP.

Croft, William (2000). *Explaining Language Change: An Evolutionary Approach*. Harlow Essex: Longman.

Cysouw, Michael (2007). Building semantic maps: The case of person marking. In Matti Miestamo, Bernhard Wälchli (Eds.), *New Challenges in Typology*, 225–248. Berlin: Mouton De Gruyter.

Dowty, David (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619.

Eisenbeiss, Sonja, Suzanne Bartke, Harald Clahsen (2005/2006). Structural and lexical case in child German: Evidence from language language-impaired and typically-developing children. *Language Acquisition*, 13, 3–32.

Fernández-Ordóñez, Ines (1999). Leísmo, laísmo, loísmo: Estado de la cuestión. In I. Bosque, V. Demonte (Eds.), *Gramática Descriptiva de la Lengua Española*, vol. I, 1319–1390. Madrid: RAE – Espasa Calpe.

Gil, David (2002). Riau Indonesian sama, a unified analysis. In U. Tadmor (Ed.), *Studies in Malay Dialects Part 3*, *Linguistic Studies of Indonesian and Other Languages in Indonesia*, vol. 50, 1–41. Jakarta: Universitas Katolik Indonesia Atma Jaya.

Gil, David (2008). How much grammar does it take to sail a boat. In Geoffrey Sampson, David Gil, Peter Trudgill (Eds.), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press.

Goldberg, Adele (1995). *A Construction Grammar Approach to Argument Structure*. Chicago: Chicago UP.

Haspelmath, Martin (1998). Does grammaticalization need reanalysis? *Studies in Language*, 20(3), 315–350.

Haspelmath, Martin (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (Ed.), *The New Psychology of Language*, vol. 2, 211–242. Mahwah, New Jersey: Lawrence Erlbaum.

Jackendoff, Ray (1990). *Semantic Structures*. Cambridge, MA: MIT Press.

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

Kulikov, Leonid, Andrej Malchukov, Peter de Swart (Eds.) (2006). *Case Systems in a Diachronic Perspective: A Typological Sketch*. Amsterdam: John Benjamins.

Levin, Beth, Malka Rappaport Hovav (2005). *Argument Realization*. Research Surveys in Linguistics. Cambridge: Cambridge University Press.

Li, Charles, Sandra Thompson (1976). Subject and topic: A new typology of language. In C.N. Li (Ed.), *Subject and Topic*, 458–489. New York: Academic Press.

Malchukov, Andrej, Andrew Spencer (Eds.) (2009). *The Oxford Handbook of Case*. Oxford: Oxford University Press.

Moy, Joanna (2006). *Word Order and Case in Models of Simulated Language Evolution*. Ph.D. thesis, University of York.

Müller, Gereon (2002). Remarks on nominal inflection in German. In Ingrid Kaufmann, Barbara Stiebels (Eds.), *More than Words: A Festschrift for Dieter Wunderlich*, 113–145. Berlin: Akademie.

Müller, Stefan (2001). Case in German – towards and HPSG analysis. In Tibor Kiss, Detmar Meurers (Eds.), *Constraint-Based Approaches to Germanic Syntax*. Stanford: CSLI.

Sperber, Dan, Deirdre Wilson (1986). *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.

Steels, Luc (1995). A self-organizing spatial vocabulary. *Artificial Life*, 2(3), 319–332.

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 Walter Daelemans (Ed.), *Proceedings 42nd Annual Meeting of the Association for Computational Linguistics*, 9–19. Barcelona.

Steels, Luc (2005). What triggers the emergence of grammar? In A. Cangelosi, C.L. Nehaniv (Eds.), *AISB'05: Proceedings of EELC'05*, 143–150. Hatfield: AISB.

Steels, Luc (Ed.) (2011a). *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Steels, Luc (2011b). A first encounter with Fluid Construction Grammar. In Luc Steels (Ed.), *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Steels, Luc (2012). Self-organization and selection in cultural language evolution. In Luc Steels (Ed.), *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins.

Steels, Luc, Jean-Christophe Baillie (2003). Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems*, 43(2-3), 163–173.

Steels, Luc, Joachim De Beule, Nicolas Neubauer (2005). Linking in Fluid Construction Grammar. In *Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC '05)*, 11–18. Brussels, Belgium.

Steels, Luc, Martin Loetzsch (2012). The grounded naming game. In Luc Steels (Ed.), *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins.

Steels, Luc, Remi van Trijp (2011). How to make construction grammar fluid and robust. In Luc Steels (Ed.), *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Steels, Luc, Remi van Trijp (to appear). Multilevel alignment maintains language systematicity. *Advances in Complex Systems*.

Steels, Luc, Remi van Trijp, Pieter Wellens (2007). Multi-level selection in the emergence of language systematicity. In Fernando Almeida e Costa, Luis M. Rocha, Ernesto Costa, Inman Harvey (Eds.), *Advances in Artificial Life (ECAL 2007)*, LNAI 4648, 421–434. Berlin: Springer.

Van Belle, William, Willy Van Langendonck (Eds.) (1996). *The Dative. Volume 1: Descriptive Studies*. Case and Grammatical Relations Across Languages. Amsterdam: John Benjamins.

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

van Trijp, Remi (2011a). A design pattern for argument structure constructions. In Luc Steels (Ed.), *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

van Trijp, Remi (2011b). Feature matrices and agreement: A case study for German case. In Luc Steels (Ed.), *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Van Valin, Robert (2004). Semantic macroroles in Role and Reference grammar. In Rolf Kailuweit, Martin Hummel (Eds.), *Semantische Rollen*, 62–82. Tübingen: Narr.

Wellens, Pieter (2011). Organizing constructions in networks. In Luc Steels (Ed.), *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Wellens, Pieter, Martin Loetzsch, Luc Steels (2008). Flexible word meaning in embodied agents. *Connection Science*, 20(2), 173–191.