Variation, competition and selection in the self-organisation of compositionality[*]

Paul Vogt
Tilburg University; ILK / Language and Information Science
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
p.a.vogt@uvt.nl

Abstract
This chapter discusses how Darwin's evolution theory can be applied to explain language evolution at a cultural level. So, rather than viewing language evolution as a process in which the users adapt biologically to learn language, languages themselves adapt to the learning abilities of individuals. Within this framework, languages evolve through variation, competition and selection. Invention and learning are identified as variation mechanisms; learnability, transmission bottlenecks and stability are pressures for competition; and optimising for success is a good selection mechanism. Rather than studying the language development in individual users, this chapter illustrates how artificial multi-agent systems equipped with these principles can self-organise a compositional language from scratch. It is argued that this model offers a good alternative to many standard approaches in linguistics.

1 Introduction

Language is possibly the most complex form of cognitive behaviours exposed by humans and may well have laid the foundation for the high level of intelligence that we generally attribute to the human species. Not surprisingly, we have great difficulties understanding exactly how we acquire and process language. One recent trend is to view languages as self-organising systems that organise on a global scale as the result of local interactions (Croft, 2002; Steels, 1997; De Boer, 2005).

This view contrasts with the idea that language has evolved into a genetically encoded Universal Grammar (Chomsky, 1956; Pinker & Bloom, 1990), which suggests that humans are innately endowed to acquire the grammatical structures of language constrained and formed by UG. One problem with this approach is that it is only concerned with the *perfect* competence of an *idealised* speaker, rather than with the performance of speakers (see, e.g., Croft, 2002, for a discussion).

---

[*] Draft version to appear in B. Wallace et al. (Eds.) (2007) *The mind, the body and the world*. Imprint.

Accounts on the self-organising nature of language (not to be confused with the nature of its speakers) are more concerned with performance and thus regard languages the way they are used, rather than the way they are represented in an idealised manner. Regarding languages as self-organising systems allows us to investigate the origins of languages (e.g. Steels, 1997) and the evolution of languages (e.g. Croft, 2000) as dynamical processes at a cultural instead of a genetic level. These aspects further suggest that it is more profitable to study language development and evolution at a population level, rather than at the individual level.

An important mechanism that has often been proposed is Darwinian selection, though not on genetic material, but on *linguemes* (Croft, 2000; Mufwene, 2002). According to this view, language users replicate parts of the language (e.g., speech sounds, words or grammar) with various modifications. As a result, a lot of variation tends to arise among which speakers have to select, either consciously or unconsciously, what variant they produce. This way, a global pattern (i.e. language) self-organises as the result of variation, competition and selection.

In the past decade or so, an increasing number of studies have simulated the origins and evolution of language computationally (see, e.g., Cangelosi & Parisi, 2002; Steels, 1997; Vogt, 2006b, for overviews). Most of these studies regard languages as self-organising systems (De Boer, 2005). In this chapter, I will briefly review a few recent studies using a computational model that simulates the emergence of compositional structures in language (Vogt, 2005a, 2005b).

Building further on these studies, the objective of this chapter is to illustrate how the interaction between variation, competition and selection can explain the emergence of compositionality. The next section will briefly discuss the idea of Darwinian selection in language evolution. I will then present and discuss the computational studies on the emergence of compositionality, after which the chapter concludes.

2 Variation, competition and selection

Darwin's theory of natural selection (Darwin, 1959) has not only found its way in biology, but also in other disciplines such as social science (Dawkins, 1976), linguistics (Croft, 2002) and even the philosophy of science (Hull, 1988). Let me briefly summarise the well known essence of Darwinian evolution. At its heart lies the principle of *variation*. Elements such as DNA molecules are replicated and during this process errors or mutations occur. These mutations cause variations in the gene pool. Some variations are well adaptive in the current environment and tend to survive; others are not so adaptive and fail to replicate further. This process is called *selection*. The reason there has to be selection is *competition*. The different genes (or replicated individual, i.e. offspring) compete for their suitedness in the environment. Some variations that occur may not be well adapted to the environment, so the organism having this variation may not survive long enough to pass over their genes.

Following Hull (1988), who has proposed the *general analysis of selection* theory for processes applicable to biological, social and conceptual development, Croft (2000) has proposed to apply these Darwinian mechanisms to explain language evolution and language change. In Croft's model, the replicators are called linguemes, which include elementary units in language such as speech sounds, syllables, morphemes, words and grammatical units. Linguemes are acquired by individuals primarily based on what they hear in their environment. Since different speakers use different variants of a lingueme or because some linguemes are transmitted with noise, a hearer will acquire a variety of linguemes. In order to produce or interpret an utterance, individuals select linguemes from this pool. Selection can – possibly unconsciously – be based on the social status of the speaker or the hearer, but may also be based on a drive to communicate effectively.

Just as each individual acquires a lingueme pool, the entire speech community forms a lingueme pool. As linguemes have been produced by a variety of speakers, the pool contains a lot of variation, from which certain linguemes are selected by the same or new speakers and possibly undergo further mutations. The evolution of this lingueme pool marks language change. A good theory on language evolution in this line of thinking will need a number of mechanisms that cause variation, competition and selection on all types of linguemes. Mechanisms causing variation include invention and mutation. Invention is typically necessary when the current language (of an individual) is insufficient for communicating something. For instance, when a new product is invented and put on the market, this may require the invention of a new word. Sounds can mutate due to noisy transmission and word-meaning mappings can also mutate due to errors in learning. Word-meanings, however, can also change by reusing an existing word to refer to a new product. Language contact is, of course, another source for variation (Mufwene, 2002). As we shall see, variations may also occur through the creative productivity of individuals through newly discovered generalisations.

Since an uncontrolled growth of the lingueme pool causes unstable communication systems to emerge, competition is required. Stability is considered to be one of the major pressures for competition (Cavalli-Sforza & Feldman, 1981), which may relate to particular pressures regulating understandability, learnability (Deacon, 1997) or social status (Croft, 2002).

To facilitate competition, one or more selection mechanisms are required. One mechanism that seems plausible is optimisation of understandability. With such a mechanism, individuals tend to select elements that have been used successfully in the past. In addition, this may be achieved by individuals aligning their communication to what they expect their speech partners will most likely understand (Pickering & Garrod, 2004).

Based on such mechanisms, De Boer (2001) has convincingly shown how a population of simulated agents equipped with a quasi-realistic model of the human vocal tract and auditory system could develop human-like vowel systems. Interactions among agents were modelled by means of imitation games, where an imitator tried to imitate the vowel produced by a speaker. If the imitated vowel was perceived by the speaker as the one it produced, the game was considered successful and the imitator would shift this vowel closer to the vowel perceived. Occasionally, the agents invented new random vowels and when two vowels came very close to one another, they were merged. As a result of the population playing many imitation games, De Boer showed that, under certain conditions, the distribution of different vowel systems that arose in different runs of the simulations were highly similar to the distribution of human vowel systems across the world's languages as reported in Maddieson (1984).

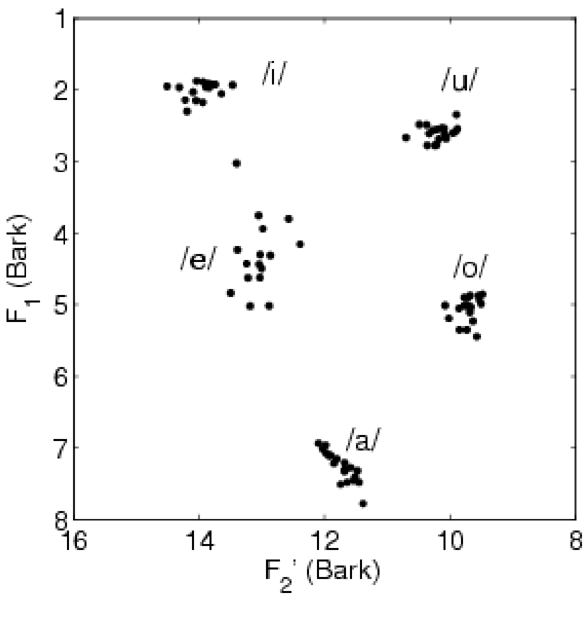

Figure 1: An example of the type of vowel systems that evolved in De Boer's simulations. The figure shows 5 clusters of vowels. In each cluster, one dot represents the vowel of one agent. Clearly, there is quite some variation in the system, though at this stage, the system is quite stable. (Reprinted with permission from De Boer, 1999.)

De Boer's model fits well with the theory of variation, competition and selection. Variation in his model was introduced in three ways: random insertion of new vowels, noise in transmission and shifting vowels toward physically heard ones. As a result of the random insertion, vowels were introduced in different areas of the vowel space. Due to the attraction of vowels (through shift), clusters were formed in different regions. Because an agent could have several vowels near one cluster, different vowels in the system competed for occupying a region in the vowel space. The individual selection mechanism, of taking the nearest neighbouring vowel to the one heard, caused vowel systems to evolve towards relatively stable systems, such as shown in Figure 1. So, the agents' tendency to imitate each other as closely as possible was the main selection mechanism in De Boer's model.

This type of approach has not only been applied to the evolution of vowel systems, but also to the evolution of lexicons (Oliphant, 1996), lexicon grounding (Steels & Vogt, 1997), syntax (Kirby, 2001) and grammar (Steels, 2005; Vogt, 2005a). In the remainder of this chapter, I will discuss how the principles of variation, competition and selection can be used to study the self-organisation of compositionality in languages.

3 The emergence of compositionality

Before presenting the computational studies, a definition of compositionality is required. *Compositionality* refers to a representation (e.g. an utterance) where the meaning of the whole is a function of the meaning of its parts and the way they are combined. An example of a compositional utterance is "green square", where the part "green" refers to the colour green and the part "square" to a rectangular shape with equal sides. Combined they form the whole meaning referring to a green square. One important aspect of compositionality is that parts can be recombined with different parts, referring to different things. For instance, "green" can be combined with "triangle" to form the utterance "green triangle". In contrast, a *holistic* expression is an expression of which the meaning of the whole is not a function of its parts. For example, there is no part in the expression "bought the farm" that refers to any part of its meaning died.

It has been shown that if an initially holistic language is transmitted iteratively over subsequent generations, this system can transform into a compositional one if the language is transmitted through a *transmission bottleneck* (i.e. learners only observe a small subset of the language), provided that learners are equipped with a mechanism to acquire compositional structures (Kirby, 2001). This is consistent with Wray's (1998) hypothesis that complex languages evolved from holistic protolanguages.

The model I will discuss is based on Kirby's model, though changed in a number of crucial ways, which allows compositionality to arise under different conditions as well.

3.1 Modelling language games

This model is based on the language game model (Steels, 1997) in combination with the iterated learning model (Kirby, 2001). It is implemented in a simulation of the Talking Heads experiment (Steels, Kaplan, McIntyre, & Van Looveren, 2002).[1] The simulation implements a multi-agent system in which the population evolves a communication system from scratch to describe coloured geometrical shapes. Each agent starts to `live' without any categories or linguistic knowledge; these are all constructed during their lifetimes by engaging in a series of interactions, called language games. It is beyond the scope of this chapter to present the model in detail, and in the following some explanations are simplified to benefit the readability without doing away with the general principles of the model. Full details are in Vogt (2005a).

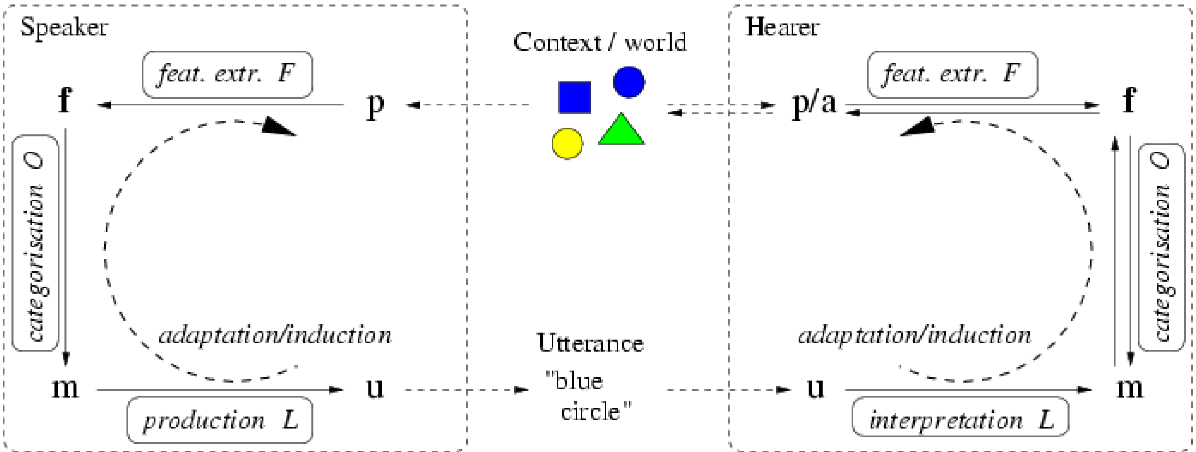

Figure 2: This figure shows a coupled semiotic square that summarises the guessing game. See the text for details.

One type of language game implemented is the *guessing game*, which is summarised in Figure 2. This game is played by two agents: a speaker and a hearer, who are presented a context that contains a given number of coloured shapes. Both agents extract perceptual features from each object and categorise these. The speaker selects one object as the topic of communication and encodes an utterance by searching his grammar for the best way to express the object. In turn, the hearer decodes the utterance by searching her grammar for the best way to parse the expression such that it is consistent with one of the objects in the context. This way, the hearer

[1] This implementation is freely available at http://www.ling.ed.ac.uk/~paulv/thsim.html.

guesses the speaker's intention and subsequently points at this object. The speaker verifies if this is the intended object and if it is, he acknowledges the hearer's success; if not, the speaker points to the topic, thus providing corrective feedback on the utterance's reference. At the end of the game (or in some cases during the game), the agents adapt their memories.

There are two types of knowledge agents acquire during their lifetimes. The first type are prototypical categories, acquired whenever the categorisation of objects fail. Objects are categorised with prototypes nearest to the objects' features representing colours and shapes. Each agent is forced to categorise each object such that its category distinguishes this object from the rest of the context. If this fails, the agent will add a new prototype to its memory for which the object's features serve as an exemplar. This categorisation scheme is called a *discrimination game* (Steels, 1997). So, early during an agent's development, the discrimination games will fail frequently, but the agent will rapidly develop an ontology that allows successful categorisation. The second type of knowledge is the grammar, which consists of rules that associate expressions with meanings (or *categories*)[2] either holistically or compositionally.

An example grammar is shown in Figure 3. In this example, rule 1 is holistic and rule 2 is compositional with terminal rules 3 and 4 as possible fillers. The grammar is constructed by invention or acquisition. If the speaker fails to encode an utterance (i.e. no rule combination matches the topic's category), he will invent a new random string either to be associated with the part of the meaning that does not match, or with the whole meaning – in which case a holistic rule is created. For instance, if the speaker of the example grammar wants to communicate about a red square categorised with (1,0,0,1), only the part (1,0,0,?) matches rule 3, so rule B → x/(?,?,?,1) is invented, where x is a random string constructed from a limited alphabet. If this string is, e.g., "toma", the speaker can now produce the utterance "redtoma" to convey the meaning of red square.

1	S → greensquare/(0,0,1,1)	0.2
2	S → A/rgb B/s	0.8
3	A → red/(1,0,0,?)	0.6
4	B → triangle/(?,?,?,0)	0.7

Figure 3: An example grammar fragment. The grammar contains rules that rewrite a non-terminal into an expression-meaning pair (1, 3 and 4) or into a compositional rule that combines different non-terminals (2). The meanings are 4-dimensional vectors where the first three dimensions relate to the RGB colour space and the

[2] Meaning in this study is represented by a category. I will use the term meaning if the category is associated with an expression.

fourth relates to the shape feature. The question marks are wild-cards. Each rule has a rule score that indicates its effectiveness in past language games.

If the hearer fails to decode the utterance or guesses the wrong referent, she will acquire one or more new rules. While doing that, she will try to generalise her language by searching for similarity patterns that allow her to form compositional rules in a usage-based fashion, such as proposed by, e.g., Lieven, Behrens, Speares, and Tomasello (2003); Tomasello (2003). As in human speech, utterances are transmitted without directly observable word boundaries, nor do agents have prior knowledge how to break up the meaning space.

If a part of the received utterance can be decoded correctly, the rest of the utterance will be associated with the rest of the meaning. For instance, suppose the hearer of the example grammar heard the utterance "redsquare", and that she knows, through corrective feedback, that the utterance refers to the object categorised as (1,0,0,1). Since the part "red" can be correctly decoded with meaning (1,0,0,?), the part "square" can be associated with meaning (?,?,?,1) resulting in the newly acquired rule B → square/(?,?,?,1). When the hearer fails to perform such an acquisition, she will see if there are any similarities between the heard utterance-meaning pair and previously heard utterance meaning-pairs stored in the agent's memory. Suppose the hearer heard "greencircle" meaning (0,0,1,0.5), then there is a regular pattern when comparing it to "greensquare" meaning (0,0,1,1), namely "green" and (0,0,1,?). This allows the agent to break up these these utterances in the parts "green", "circle" and "square" with corresponding meanings, forming the terminal rules A → green/(0,0,1,?), B → circle/(?,?,?,0.5) and B → square/(?,?,?,1). If the corresponding compositional rule, such as rule 2 in Figure 3, does not yet exists, this will also be constructed. If this acquisition mechanism also fails, the hearer incorporates the utterance-meaning pair holistically.

If new knowledge is acquired, old knowledge is not thrown away (which is the case in Kirby's, 2001, model). This is important, because it allows different variants to compete with each other. In principle, each meaning can be associated with different utterances and each utterance can be associated with multiple meanings. Also compositional rules may compete with each other. Competition is regulated primarily with the rule score σ_r. Every time a rule is used to encode or decode an utterance successfully, its score is increased by

$$\sigma_r = \eta \cdot \sigma_r + 1 - \eta \qquad (1)$$

where $\eta = 0{:}9$ is a constant learning parameter. At the same time, competing rules (i.e. those rules that either match the meaning or the utterance) are laterally inhibited by

$$\sigma_r = \eta \cdot \sigma_r \qquad (2)$$

If a rule is used unsuccessfully, its score is decreased by this same equation. Given that initial scores are between 0 and 1, these updates make that the rule score is always a value between 0 and 1.

When a speaker or hearer has to select between two or more competing compositions of rules when decoding or encoding an utterance, they will always select the one with the highest combined score. The combined score is the product of all scores used in a composition. If the utterance is formed from a holistic rule, the composition contains one rule, otherwise it contains three rules.[3] This way, selection is biased towards holistic rules. However, a compositional rule can be used in novel situations (e.g., to talk about a red circle without ever having talked or heard this before). Moreover, compositional rules can be used in more situations than a single holistic rule. When this occurs frequently enough, the score of the compositional rule may become so high that a combination of scores using this rule can exceed the score of a holistic rule, so that the holistic rule would loose the competition.

Note, by the way, that applying a compositional rule in a novel situation does not only yield a new variation in the individual's language, but may also introduce a new variation in the global language. Hence, learnt generalisations may provide new variations that could compete with other variations.

All simulations discussed below were carried out with a population of which half were `adult' agents and the other half were `children'. After a given number of language games, all adults were removed, the children became adults and new children were added to the simulation. This type of population dynamics is similar to the *iterated learning model* (Kirby, 2001), which is a standard model for studying language evolution computationally.

The world in all experiments presented contained 120 different objects (10 shapes times 12 colours) and each guessing game was played in a context of 8 objects randomly sampled from the world. Each different experiment was replicated 10 times (i.e. 10 runs) with different random seeds.

---

[3] In the current model, compositions are formed from two constituents at most, thus using three rules (the compositional rule and the two terminal rules).

3.2 Transmission bottlenecks

Before reviewing some of the results, it is important to note that complex dynamical systems such as the one presented here tend to evolve toward *stable states* (Cavalli-Sforza & Feldman, 1981). As mentioned, it was shown by Kirby (2001) that compositional systems can emerge from holistic ones, provided children are equipped with acquisition mechanisms to discover and acquire compositional rules and provided the language is transmitted through a bottleneck. This transmission bottleneck means that children only observe a small subset of the adults' language. This transition from holistic to compositional languages can be understood by realising that only compositional languages can be transmitted stably through a bottleneck. If a language is holistic, learning from observing only a subset requires the invention of new utterances for the elements not observed.

Consider, for example, this simple language with four holistic utterance-meaning pairs: *toma-[red,square]*, *tupa-[green,triangle]*, *bulo-[green,square]* and *rino-[red,triangle]*. If you only see the first three instances, you need to create a new utterance for meaning *[red,triangle]* when you wish to convey it. However, if the language is compositionally structured, such as *toma-[red,square]*, *bulo-[green,triangle]*, *buma-[green,square]* and *tolo-[red,triangle]*, then learning from the first three instances allows you to recreate the entire language. Hence, compositional languages are stable under the transmission of a bottleneck, whereas holistic languages are not. However, when the holistic language has some fortuitous pattern in some utterance-meaning pairs, these may be extracted and have a higher chance to survive the bottleneck. This, then increases the chance that a new variation that has a regular pattern may enter the global language by combining extracted patterns, thus increasing the amount of compositionality in the language.

This is all very well, but using the model explained above and with a population size of 1 adult and 1 child, and considering only *vertical transmission* (which are the same conditions as in Kirby, 2001), it has been shown that compositionality can emerge stably without a bottleneck (Vogt, 2005a). – Vertical transmission means that the language is transmitted from one generation to the next. In such a protocol all speakers are adults and all hearers are children. – Why is that the case? First, recall that agents can form new compositional rules f they find a similarity in two utterance-meaning pairs. The chance that this happens in this model is quite large. Second, once compositional rules exist, they apply to more situations than a holistic rule and are therefore used more frequently. If rules are used more frequently, their scores tend to increase more strongly than those of rules used less frequently. So, the competition between holistic rules and compositional ones will be won by the compositional ones. In Kirby's model there is no competition between different rules, because his learning mechanism only allows one-to-one mappings between utterances and meanings. In that case, there is no need to form compositional rules when there is no bottleneck. Some will emerge, but holistic rules have a good chance to survive.

When, however, the population size is increased from 2 to 6 agents (3 adults and 3 children), transmission through a bottleneck is required for stable compositional systems to emerge (Fig. 4). This is because there are more agents in each generation and agents of one generation do not communicate with each other in this vertical transmission protocol (all speakers are adults and all hearers are children), so there arises much ambiguity in the input to the children, which – in this case – leads to instability of compositionality (note that communicative success is hardly affected). Moreover, it has been shown that in this model the incremental development of categories can cause words to be associated with very broad categories in early stages of development, leading to overextensions and eventually to dramatic changes of a word's meaning (leading to meaning drift) when they are narrowed down, which also leads to unstable solutions (Vogt, 2006c). Since ambiguities or meaning drift have more impact on compositional languages than on holistic ones[4], the competition is won by holistic rules. So, in the case with larger populations and vertical transmission with a bottleneck, variation is not good for compositionality as it leads to unstable systems. What seems to be lacking is a good competition pressure.

The transmission bottleneck turns out to be a good pressure, because with a bottleneck, agents will tend to use compositional rules more frequently when they need to communicate about previously unseen meanings. Consequently, such rules will be reinforced and thus tend to win the competition. In a way, the bottleneck tightens the competition, which triggers the selection of more compact and combinatorial languages.

Up to now, I discussed only simulations with a vertical transmission of language. However, children in our society do not start speaking when they are grown up; they start speaking from the age of one or even before. In a *horizontal transmission* scheme, where both speakers and hearers are randomly selected from the entire population (including both adults and children), communication goes in all directions within the population.[5] In

---

[4] Imagine what will happen if the meaning of "red" will change into green in comparison to changing the meaning of "bought the farm" to life. The former change will definitely have a larger impact on the language as a whole than the latter.

[5] This definition of horizontal transmission is consistent with the one used in epidemiology, where it relates to a transmission between all members of a population irrespective of their

such a case, compositionality does evolve to a stable state, even in the absence of a transmission bottleneck (see, Fig 4, Vogt, 2005b).

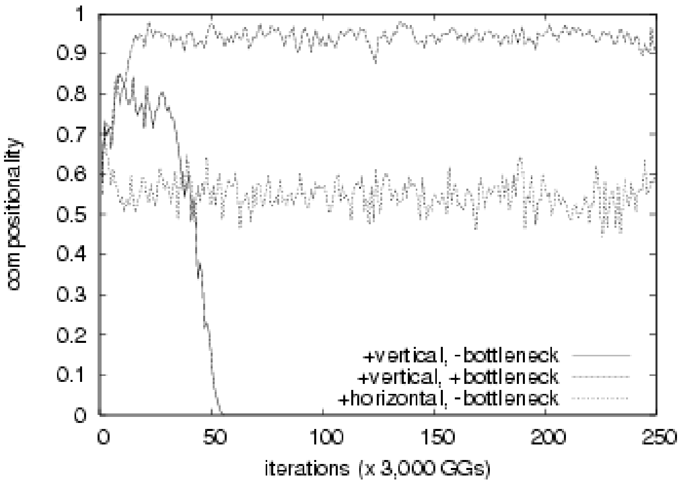

Figure 4: Compositionality as a function of the number of iterations (or generations). Compositionality is measured as the proportion of utterances produced or interpreted using a compositional rule, rather than a holistic one. In each iteration 3 adults and 3 children engaged in 3,000 guessing games, after which the adults are replaced by the children and new children entered the population. Results are provided for the cases where there was no bottleneck (solid line), with a bottleneck (dashed line) – both with vertical transmission – and one without a bottleneck, but with horizontal transmission (dotted line). This figure is adapted from Vogt (2005b).

The reason why in the absence of a transmission bottleneck compositionality evolves to a stable state is that for horizontal transmission children sometimes need to speak about things they never encountered before. They, thus, face the consequences of the bottleneck, even though this bottleneck is not imposed by the experimenter (which was necessary in the case of vertical transmission). As explained before, when the child needs to communicate about a meaning it did not see before (e.g., meaning *[red,triangle]*), while it did learn the utterance-meaning pairs of this meaning's parts (e.g., it learnt from hearing *toma-[red,square]*, *bulo-[green,triangle]*, *buma-[green,square]*), the child does not need to invent a new word, but can select a compositional rule to express the meaning. This *implicit bottleneck* effect is a natural consequence of the normal development and interactive communication of the child. Again, when compositional rules are used more frequently, they are reinforced more strongly (at least when they are used successfully), and consequently more likely to be reselected, thus forming a positive feedback loop.

---

generation. This is different from the definition given by Cavalli-Sforza and Feldman (1981) who consider horizontal transmission only within one generation.

3.3 Population size effects

So far, all simulations were carried out with a very small population size (starting with only 2 agents – 1 per generation – and then with 6 agents). In Vogt (2005c), it was investigated what happens if we increase the population size. Interestingly, after a first decline for populations up to 30 agents, the level of compositionality increases up to around 95% when the population further increases to 100 agents – at least for those runs that yielded compositionality higher than 50% (Fig. 5, top). Interestingly, only those

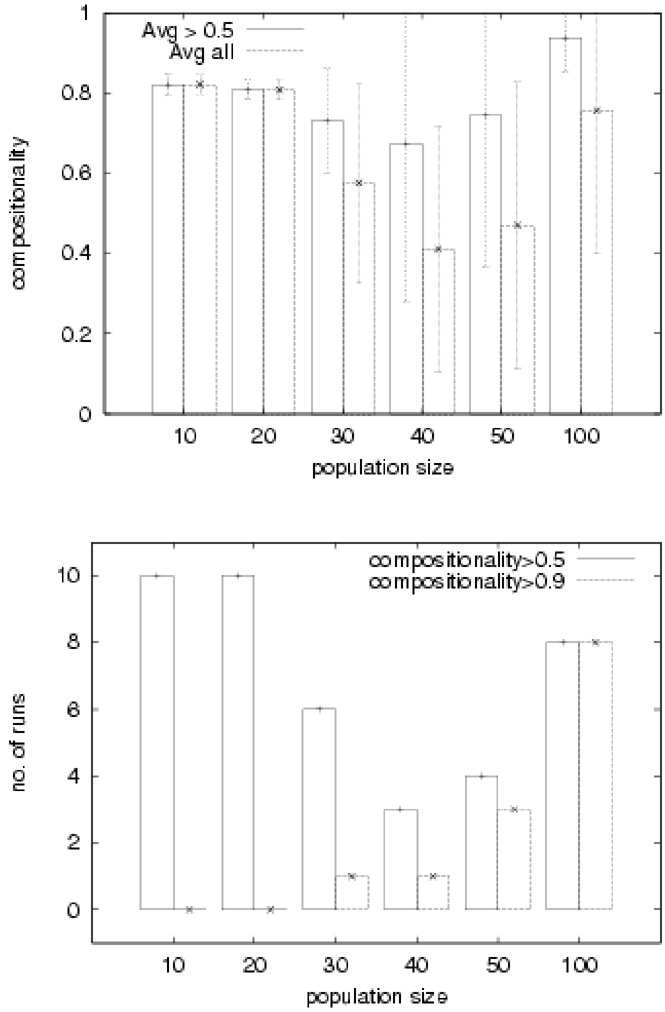

Figure 5: (top) The level of compositionality reached at the end of simulations with different population sizes. The solid boxes give the results averaged for only those runs that yielded compositionality higher than 0.5 and the dashed boxes give the averages for all 10 different runs per condition. The error-bars indicate standard deviations. (bottom) This graph shows the number of runs out of 10 that yielded compositionality higher than 0.5 and those higher than 0.9.

simulations with a population size larger than 30, yielded compositionality of levels higher than 90%, while the simulations with 100 agents performed best (Fig. 5, bottom). More precisely, for a population size of 100, 7 out of 10 different runs yielded compositionality higher than 0.9.

The reason for this increase in compositionality is to be sought in the increased level of variation and competition. As there are more agents in the population, each starting without any linguistic knowledge, more new words will be invented. This will increase the chance that different utterance-meaning pairs will find a similarity that allows agents to break up utterances and form compositional rules. As a consequence, the chance is also increased that good (i.e. effective) compositions are formed. However, having a larger population also makes it is harder to achieve communicative success (the level of communicative success increases slower for larger populations). There is more competition (due to more variation), but it is also harder for all agents to understand each other agent. This competition makes the selection more crucial. It appears that compositionality does help in this, as in consecutive generations learning appears to become more and more effective (i.e. higher levels of success are reached, while similar levels are reached faster).[6] So, it seems that the language evolves to become better learnable; a conclusion that was also reached by Kirby (2001), though for slightly different reasons.

3.4 Population dynamics

The learnability of the language becomes more apparent when looking at the population dynamics of the system, which has been studied in Vogt (2006a). This study started from the observation in Vogt (2005b) that the evolution across generations does not seem to change a lot when the language is transmitted horizontally. So, to what extent does it not change and what is the added value of a population turnover – if any? To study this question, a comparison was made between a simulation that contained only one generation and one in which, halfway during the simulation, half of the population was replaced by children. So, in effect in the first case there was only one iteration and in the second case there were two. The total population size in these simulations were 50 agents. The results, reported in Vogt (2006a), are summarised in Fig. 6.

The graphs show that for compositionality (top) and communicative success (middle) there are similar evolutions for both conditions, though the simulation considering two generations (right) showed a short discontinuity when the population was changed. The two bottom graphs show the relative

[6] A more extensive study on population size effects has recently been published in Vogt (2007).

frequencies with which different rule types are used and need a bit more explanation.

The agents can form holistic rules (rule type I) and compositional rules by breaking apart the 4-dimensional conceptual space in different ways. They can form rules by combining colour and shape (rule types IV and V), but they can also combine, for instance, the red component of the RGB colour space with the blue and green components together with shape (i.e. red vs. blue, green and shape – rule types II and III).[7] The 10 other ways to combine the different dimensions of the meaning space do occur, but with negligible frequencies.

Figure 6 (bottom left) shows that initially, there is a lot of competition among the different rule types, but that around game 500,000 a more or less stable system has developed. In this system, there is a relatively high incidence of rule types IV and V, but rule types I and II also occur quite frequently. So, this language has stabilised more or less in a sub-optimal system in the sense that the grammar is not in the most compact form, which requires the language to be formed of rules combining colours with shapes only (i.e. rule types IV and V). Nevertheless, communicative success is quite high.

When after the initial competition the adults are removed and new children are introduced, the language changes rapidly and rule types IV and V start to dominate while others tend to die out (Fig. 6, bottom right). This happens because the new agents rapidly learn these rules and tend to reuse them in cases where they have to communicate about previously unseen meanings due to the implicit bottleneck. The reason for the fact that these rules are rapidly learnt and used is that they apply in all possible situations, whereas all other rule types are less optimal. (Combining each colour with each shape relates to all objects in the world, but combining, e.g. each value of the red component with all possible values in the other dimensions may give descriptions of objects that do not exist in the virtual world.) This allows the children to form new combinations of words that can effectively convey their referents. So, where the older generation may express a number of meanings using holistic rules or compositional rules of type II and III, children tend to introduce novel combinations (i.e. variations) in the language using mainly rule type IV and V.

[7] For each way of breaking up the 4 dimensional space, there are two types indicating word order: e.g., IV is colour first and V is shape first.

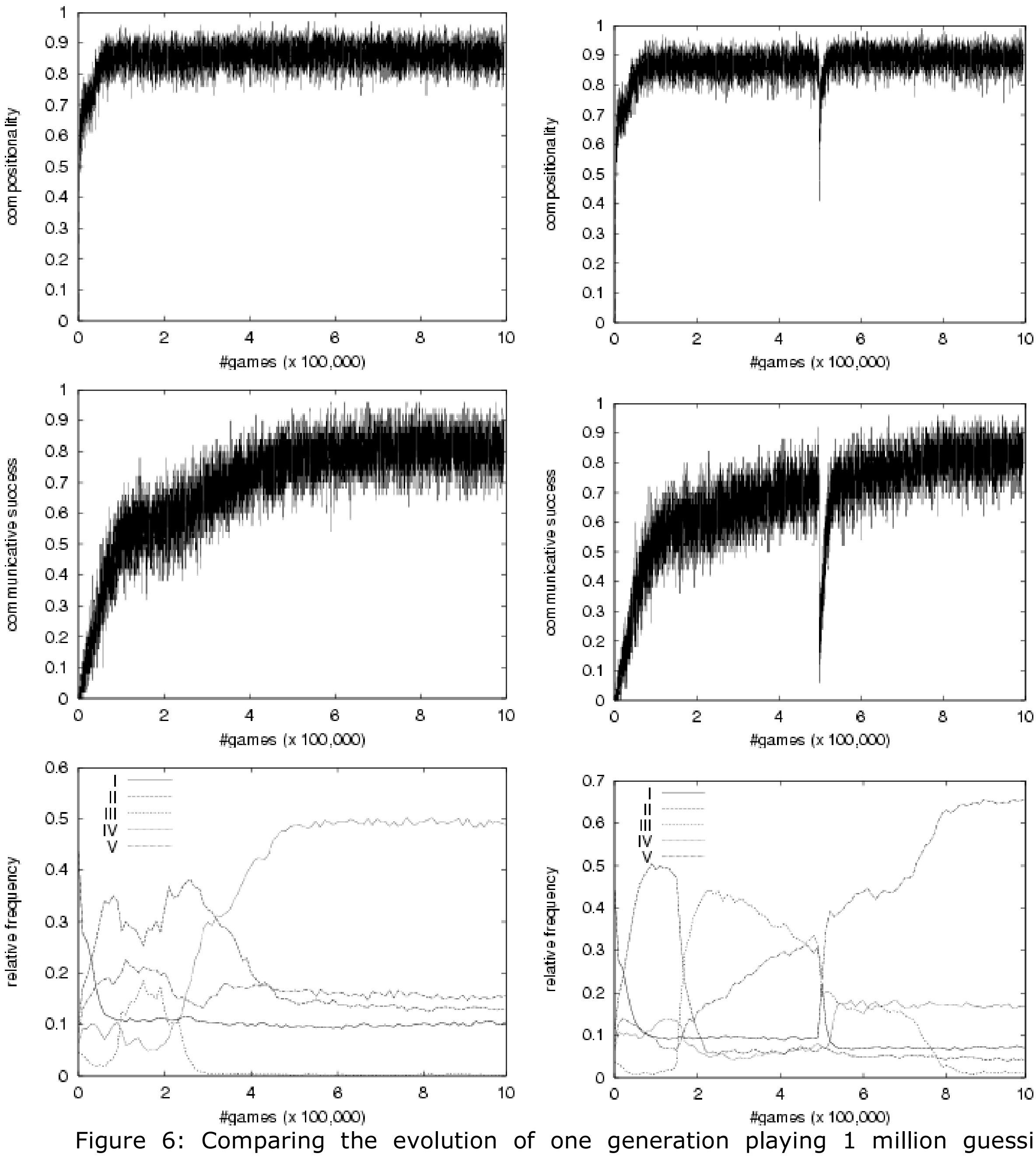

Figure 6: Comparing the evolution of one generation playing 1 million guessing games (left) with two generations playing a total of 1 million guessing games in two iterations (right). The graphs show compositionality (top), communicative success (middle) and rule frequencies (bottom). All results relate to one simulation run for both conditions.

The chance that such new variations become rapidly successful is quite high, because part of the older generation will have acquired similar rules, even though they may not actively use them for certain situations. As a result, the uses of these rules are reinforced by all agents (young and old), so they tend to be reused more frequently. Simulations over more than two generations have shown that eventually the language evolves to an almost entire use of rule types IV and V (Vogt, 2006a). Moreover, similar results have been achieved in simulations where the population ow is more natural, in contrast to the presented catastrophic population change at the end of each iteration (Tamariz & Vogt, in prep.).

4 Implications

This chapter has reviewed a number of recent studies regarding the evolution of compositional structures in language using computer simulations (Vogt, 2005a, 2005b, 2005c, 2006a, 2006c). This review places these studies in the context of a neo-Darwinian usage-based approach to language evolution, similar to those proposed by, e.g., Croft (2000); Mufwene (2002); Tomasello (2003). It shows that the neo-Darwinian approach to language evolution is a fruitful approach that can explain, at least, the evolution of communicative successful languages which have a limited level of compositional in a population. Crucial to this approach is to study language development and use in populations, rather than to study the language competence of individuals as is done in the Chomskyan tradition and many (if not most) psychological studies on language development.

The model with which the simulations were carried out implement learning mechanisms that allow the development of a constructive grammar similar to those described by Lieven et al. (2003); Tomasello (2003) and which are general enough to count as a general learning mechanism that need not have evolved specifically for language. Although all individuals in all generations started with these learning mechanisms, transitions from initially holistic languages to well structured compositional languages have been shown. Moreover, simulations have shown that the languages themselves evolve to become learnable, rather than that the individual language users evolve to learn the languages as an innate theory such as proposed by Pinker and Bloom (1990) would predict.

The simulations discussed do not provide unequivocal evidence that language has evolved this way, however, they do illustrate a clear alternative to the nativist theories advocated by, e.g., Chomsky (1956); Pinker and Bloom (1990). Assuming the origins of language coincides with the appearance of Homo Sapiens, language arose some 250,000 years ago and given the – on an evolutionary time scale – short time it took for modern languages to have evolved from their precursors, cultural evolution seems a more likely candidate to explain a transition from holistic protolanguages to modern languages. Certainly, a number of major biological adaptations, such as adaptations relating to our speech organ (Fitch, 2000) or the emergence of Theory of Mind related issues (Tomasello, 2003), must have facilitated language evolution, but whether there were language specific adaptations as proposed by Chomsky or Pinker and Bloom is questionable. For instance, the ability to acquire compositional structures (e.g., predicate logic) could have evolved for more general cognitive abilities such as vision and can even be found in certain other species (Hurford, 2004). In addition, Parker (2006) has argued that the same could even hold for what Hauser, Chomsky, and Fitch (2002) have called the *narrow language faculty* that has recursion as its hallmark and which they hypothesise is what makes homo sapiens unique regarding their language abilities.

So, I argue that having non-domain specific learning mechanisms which try to extract regular patters from the speech input in relation to regular patterns in their meanings, such as proposed by Lieven et al. (2003); Tomasello (2003), could be sufficient to explain the transition from holistic languages to compositional or even syntactic recursive languages (Kirby, 2002). One reason why such an explanation is to be favoured is that such a fast process is possibly more easy to achieve than a slow biological evolution. However, such a cultural cumulation of complexity in language (and perhaps culture in general) seems only to be possible if (see also, Vogt, 2006a, for a related discussion):

1. language is transmitted repeatedly from one generation to the next,
2. there is sufficient variation in the language for novel traits to be discovered,
3. language is sufficiently well-structured so that the cost of learning is lower,
4. there is room for cognitive development, and
5. there is an ecological niche that attracts further development.

As the final set of experiments showed, the structure of language can evolve towards a local maximum, which may be sufficient for communication, but not for cumulating knowledge. The new generation introduces new variation that triggers the system to get out of the local maximum (in a way, this is similar to simulated annealing used in many AI systems to get out of local maximums). In addition, these children face an implicit bottleneck that serves as a competition pressure for selecting compositional structures. For complexity to cumulate, the language must also evolve so that it becomes easier to learn, which frees time for the population to create new structures. Though in the simulations discussed this only happens early in evolution (i.e. first few iterations), decreasing the cost of learning will be a prerequisite to understand the explosive growth in cultural knowledge and possibly also for explaining the complexification of language (see also, Boyd & Richerson, 2005).[8]

Of course, in order to increase the complexity of language, there must be enough room for cognitive development. In the current set up, agents are restricted to form only two word sentences. However, if the cognitive architecture would allow more complex sentences, such sentences would evolve, though maybe these will not be very efficient. In addition, the language development could face another limit, namely that of the environment. In the model the environment only provides four perceptual features with a limited number of objects. If the language has evolved such that it reects those features and those objects, there is no possibility to become more complex as the language can then describe all possible aspects of the environment. So, there must be an ecological niche that attracts further development. Of course, our natural environment is far more complex than that of the model, but it may be that our contemporary language surpasses the relevant structure of the ecological niche from, say, 250,000 years ago. However, humans not only observe the environment (which is what the agents in the model do), but also change it so there is ecological or cultural niche construction (Laland, Odling-Smee, & Feldman, 2001). Niche construction (be it ecological or cultural) can change the niche such that new targets arise to which our culture or language evolves to.

5 Conclusions

In this chapter, I have discussed a Darwinian approach to explain language evolution. This approach was not on a biological level, but on a cultural language level. In short, I have explained how variation, competition and selection can account for language change and I have reviewed a number of recent computational studies to show how these mechanisms can explain a self-organisation of compositional structures in languages.
The studies suggest that variation, competition and selection can explain how compositionality in languages can arise. Variation is crucial for setting the right conditions for competition and selection. Too much variation can harm the system if this leads to too much ambiguity. However, more variation can also lead to better performances if it increases the chance of finding good solutions. In addition, the creative process of recombining learnt parts when the consequence of transmission bottlenecks occur can also be seen as a vital selection mechanism.

---

[8] If languages become easier to learn, one may wonder if talking about complexification is justified.

Competition pressures are required to allow the system self-organise toward a (quasi) stable state. (I use the term quasi, because the system keeps on evolving, i.e. changing, even though the level of compositionality or communicative success remains stable.) Language stability, transmission bottlenecks and learnability have been identified as possible pressures for competition. Selection processes are important ingredients that should serve the purpose of competition. In the current model, optimisation has proved to be a viable selection mechanism for stable languages to develop.

Concluding, Darwin's theory on variation, competition and selection can well be applied to explain how languages can evolve. It offers a powerful alternative to more traditional approaches taken in linguistics. Finally, the computational model presented is a starting point to investigate how (grammatically) more complex languages can evolve.

Acknowledgements

Most of the work presented in this chapter was carried out at the Language Evolution and Computation Research Unit of the University of Edinburgh. I am grateful to all of its members for their comments and suggestions on this work. The work has been financed by the EC through a Marie Curie Fellowship and by the Netherlands Organisation for Scientific Research (NWO) through a VENI grant.

References

Boyd, R., & Richerson, P. (2005). *The origin and evolution of cultures.* Oxford: Oxford University Press.
Cangelosi, A., & Parisi, D. (Eds.). (2002). *Simulating the evolution of language*. London: Springer.
Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach.* Princeton, NJ: Princeton University Press.
Chomsky, N. (1956). Three models for the description of language*. IRE Transactions on Information Theory IT, 2(3)*, 13-54.
Croft, W. (2000). *Explaining language change: An evolutionary approach.* Harlow: Longman.
Croft, W. (2002). The Darwinization of linguistics. *Selection, 3*, 75-91.
Darwin, C. (1959). *The origin of species.* London: John Murray.
Dawkins, R. (1976). *The selfish gene.* Oxford: Oxford University Press.
Deacon, T. (1997). *The symbolic species.* New York, NY.: W. Norton and Co.
De Boer, B. (1999). *Self-organisation in vowel systems.* Unpublished doctoral dissertation, Vrije Universiteit Brussels.

De Boer, B. (2001). *The origins of vowel systems.* Oxford: Oxford University Press.

De Boer, B. (2005). Self-organisation in language. In C. Hemelrijk (Ed.), *Self-organisation and evolution of biological and social systems.* Cambridge UK: Cambridge University Press.

Fitch, W. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences, 4 (7),* 258-267.

Hauser, M., Chomsky, N., & Fitch, W. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science, 298 (5598),* 1569.

Hull, D. (1988). *Science as a process.* Chicago: Chicago University Press.

Hurford, J. (2004). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences, 26 (3)*, 261-283.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5(2),* 102-110.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge University Press.

Laland, K., Odling-Smee, J., & Feldman, M. (2001). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences, 23 (01),* 131-146.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: a usage-based approach. *Journal of Child Language, 30(2),* 333-370.

Maddieson, I. (1984). *Patterns of sounds.* Cambridge: Cambridge University Press.

Mufwene, S. (2002). Competition and selection in language evolution. *Selection, 3,* 46-56.

Oliphant, M. (1996). The dilemma of saussurean communication. *Biosystems, 1-2(37)*, 31-38.

Parker, A. R. (2006). Evolving the narrow language faculty: Was recursion the pivotal step? In A. Cangelosi, A. Smith, & K. Smith (Eds.), *The evolution of language: Proceedings of the 6th international conference on the evolution of language.* World Scientific Press.

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169-226.

Pinker, S., & Bloom, P. (1990). Natural language and natural selecion. *Behavioral and brain sciences, 13,* 707-789.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication, 1(1),* 1-34.

Steels, L. (2005). The emergence and evolution of linguistic structure: From lexical to grammatical communication systems. *Connection Science, 17 (3-4),* 213-230.

Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language*. Oxford, UK: Oxford University Press.

Steels, L., & Vogt, P. (1997). Grounding adaptive language games in robotic agents. In C. Husbands & I. Harvey (Eds.), *Proceedings of the fourth European conference on artificial life.* Cambridge Ma. and London: MIT Press.

Tamariz, M., & Vogt, P. (in prep.). A co-evolutionary framework for cultural transmission: Interactions between public and mental representations. *In preparation.*

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

Vogt, P. (2005a). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence, 167(1-2),* 206-242.

Vogt, P. (2005b). On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior, 13(4)*, 325-346.

Vogt, P. (2005c). Stability conditions in the evolution of compositional languages: issues in scaling population sizes. In P. Bourgine, F. Képès, & M. Schoenauer (Eds.), *Proceedings of the European conference on complex systems, ECCS'05*.

Vogt, P. (2006a). Cumulative cultural evolution: Can we ever learn more? In S. Nolfi et al. (Eds.), *From animals to animats 9: Proceedings of the ninth international conference on simulation of adaptive behaviour.* Berlin: Springer Verlag.

Vogt, P. (2006b). Language evolution and robotics: Issues in symbol grounding and language acquisition. In A. Loula, R. Gudwin, & J. Queiroz (Eds.), *Artificial cognition systems.* Idea Group.

Vogt, P. (2006c). Overextensions and the emergence of compositionality. In A. Cangelosi, A. Smith, & K. Smith (Eds.), *The evolution of language: Proceedings of the 6th international conference on the evolution of language.* World Scientific Press.

Vogt, P. (2007), Group size effects on the emergence of compositional structures in language, In F. Almeida e Costa, L. M. Rocha, E. Costa, I. Harvey & A. Coutinho (Eds.) *Proceedings of the European Conference on Artificial Life, ECAL 2007.* Berlin: Springer Verlag.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication, 18*, 47-67.