On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children

Paul Vogt[1,2]

[1]Language Evolution and Computation Research Unit
School of Philosophy, Psychology and Language Sciences
University of Edinburgh, 40 George Square, Edinburgh, Scotland.
[2]Induction of Linguistic Knowledge / Computational Linguistics
Tilburg University, P. O. Box 90153, 5000 LE Tilburg, The Netherlands

paulv@ling.ed.ac.uk

This paper will appear in a special issue on Language Evolution and Acquisition of *Adaptive Behavior.*

Abstract

This paper investigates the productive creativity of children in a computational model on the emergence and evolution of compositional structures in language. In previous models it was shown that compositional structures can emerge in language when the language is transmitted from one generation to the next through a *transmission bottleneck*. Due to the fact that in these models language is transmitted only in a vertical direction where adults only speak to children and children only listen, this bottleneck needs to be imposed by the experimenter. In the current study, this bottleneck is removed and instead of having a vertical transmission of language, the language is – in most simulations – transmitted horizontally (i.e. any agent can speak to any other agent). It is shown that such a horizontal transmission scenario does not need an externally imposed bottleneck, because the children face an implicit bottleneck when they start speaking early in life. The model is compared with the recent development of Nicaraguan Sign Language, where it is observed that children are a driving force for inventing grammatical (or compositional) structures, possibly due to a sparseness of input (i.e. an implicit bottleneck). The results show that in the studied model children are indeed the creative driving force for the emergence and stable evolution of compositional languages, thus suggesting that this implicit bottleneck may – in part – explain why children are so typically good at acquiring language and, moreover, why they may have been the driving force for the emergence of grammar in language.

1 Introduction

It does not often occur that modern humans can witness the birth of a new language. Most modern languages have emerged ages ago, though they still keep on changing. Recent exceptions are Pidgin and Creole languages and Nicaraguan Sign Language. Pidgin and Creoles languages are languages that emerged through extensive contact between two or more language cultures, e.g., as the result of colonisation (Sebba, 1997). Nicaraguan Sign Language is a recently emerged sign language among deaf Nicaraguans (Kegl and Iwata, 1989). Although Pidgin and Creole languages mostly emerged during an era in which scientists did not observe the birth of languages, Nicaraguan Sign Language (NSL) has emerged over the past 25 years and has been well studied (Senghas et al., 2004).

During the late 1970s deaf Nicaraguan children and adults were brought in contact with each other in specialised schools. Before that, deaf people only tended to communicate with their families using *home-signs*: a communication system with basic gestures, but with some rudimentary elements of language

(Goldin-Meadow, 1982). However, over the years, when new children continued to enter the deaf community schools, NSL developed into a full fledged language with many – if not all – aspects of modern spoken languages (Senghas and Coppola, 2001).

Unlike most other sign languages, such as American Sign Language, NSL has not been designed by experts; instead, NSL emerged spontaneously through the interactions among the members of the community. Senghas and colleagues (Senghas and Coppola, 2001; Senghas et al., 2004) have argued that children rather than adults have been the creative force for the increasing complexity of grammar in NSL. So, although children are known to acquire languages more easily than adults, they also appear to be more creative in adding complexity to languages than adults.

Senghas and Coppola (2001) have shown that NSL had become more complex over several *cohorts* (or generations) of children that entered the community. They have argued that this happened due to the lack of rich input children received. If the input to a child is rich enough, this child may be able to acquire the entire language they observe without any need for added complexity. However, during the initial stages of the emerging NSL, the input was not very rich. So – according to Senghas and Coppola (2001) – in order to explain the increasing complexity of NSL, the children acquiring the language must have invented added complexity.

Recent computational models on the emergence and evolution of complex languages have shown that if languages are acquired by subsequent generations of learners who receive only a limited amount of the language from their adult generations, initially unstructured (*holistic*) languages evolve into structured (*compositional*) languages (Brighton, 2002; Kirby, 2001; Smith et al., 2003; Vogt, 2005a). However, in these models only adults talk and children only listen, thus children only learn from the adults' input. It is not until the children become adults themselves, that they start to speak to newly introduced children. This purely *vertical* transmission of language is rather unrealistic. Moreover, in these models the sparseness of the input was enforced by the experimenter, which is also not realistic. As a consequence, it is unclear whether in these models the children are really the creative driving force or not; even though children are the only ones in these models who can create new compositional structures.

A more realistic model would allow children to speak to other children, and to adults, without a sparseness of the input being explicitly imposed. The present study investigates whether or not a compositional language (i.e. a language with some grammatical structure) can emerge from an initially unstructured communication system in a computational model in which there is a generational flow of adults and children who communicate in various directions, without directly imposing sparse input. The model, introduced in Vogt (2005a), implements a simulation of the Talking Heads experiment (Steels et al., 2002) combined with the *iterated learning model* introduced by Brighton (2002); Kirby (2001). The objective of this study is to show that children are, indeed, a creative driving force for the emergence of complex structures in languages based on an implicitly observed sparseness in the input.

The next section provides some further background on existing models regarding the emergence and evolution of complex languages. The model is then presented in Section 3. Section 4 presents experimental results. The results are discussion in Section 5 and the paper concludes in Section 6.

2 Compositionality, evolution and bottlenecks

Before starting the discussion of computational models that investigate the emergence of compositionality in languages, the notion of *compositionality* needs to be defined. Following Frege (1892), a representation is compositional *when the meaning of the whole is a function of the meaning of its parts and they way in which they are combined*. For instance, the phrase "kick the ball" combines the meanings of "kick" and "the ball". The meaning of "kick" can also be combined with the meaning of "the dog" to form "kick the dog". In contrast, combining "kick" with "the bucket" forms the *holistic* phrase (or holophrase) "kick the bucket", which in English means `to die`; hence the meaning of "kick" is no longer a part of the whole meaning, nor is any other part.

type	Grammar(t)	Utterances	Grammar(t+1)	Utterances
holistic	`S->ac/00`	ac-00	`S->ac/00`	ac-00
	`S->ef/10`	ef-10	`S->ef/10`	ef-10
	`S->bc/01`	bd-11	`S->bd/11`	gh-11
	`S->bd/11`		`S->gh/01`	
compositional	`S->ac/00`	ac-00	`S->A B`	ac-00
	`S->bc/10`	bc-10	`A->a/0`	bc-10
	`S->ad/01`	ad-01	`A->b/1`	bd-11
	`S->bd/11`		`B->c/0`	
			`B->d/1`	

Figure 1: Suppose a language contains 4 meanings `00, 11, 10` and `01`. The upper language is holistic and the grammar at time t contains unstructured mappings from these four meanings to signals (the strings ac, ef, bc and bd). When this language is transmitted through a bottleneck, the next generation only observes a subset of this grammar. Suppose the next generation only observes the three utterances given in the third column, then the next generation can only acquire mappings based on these observations and if an individual from this generation wishes to communicate about the previously unseen meaning (i.e. `01` in the example), this individual has to invent a new word to convey this meaning (gh in the example). This yields the grammar at time $t+1$. When the next generation then observes another set of utterances, such as the ones in the final column), the language is bound to change even more. Hence, holistic languages cannot be maintained through a bottleneck. Now, if the language contains a regular compositional structure as shown in the bottom part of the table, the grammar at time t can be maintained. If an individual of generation $t+1$ only observes three utterances, then – provided it has a means to break up utterances based on some regularity in the signal space and the meaning space – this individual is able to acquire the compositional grammar at $t+1$. Using this grammar, this individual is capable of reproducing the entire grammar of generation t. So, compositional languages can be maintained through a bottleneck.

In essence, holophrases are unanalysed utterances, whereas compositional utterances are analysed (Wray, 1998). Wray has hypothesised that during the course of evolution, language first emerged as a holistic protolanguage, after which it gradually developed into a compositional language. Although it is debatable whether protolanguage was holistic or consisted of unstructured multiple word utterances, see, e.g., Bickerton (1990), it is widely accepted that protolanguages were precursors to modern languages, see, e.g., Jackendoff (1999). So, before languages evolved in the structured systems of modern languages, they are thought to have gone through a stage where languages were less structured. Furthermore, there is some evidence that the processing of holophrases occur in first language acquisition (Peters, 1983; Lieven et al., 2003), the same seems to hold for second language acquisition, adult native language and aphasia (Wray, 1998).

Brighton (2002); Kirby (2001); Smith et al. (2003); Vogt (2005a) and others have shown how compositional structures in languages can emerge from initially holistic languages over generations of adaptive agents by transmitting the language through a *bottleneck*. This bottleneck refers to a situation where the input to language learners is sparse, i.e. they learn the whole language while observing only a part of the language. The explanation for this language change is that holistic languages cannot be transmitted stably over generations through a bottleneck. Compositional languages, on the other hand, allow the population to maintain a relatively stable and expressive language. The reason why compositional languages can be maintained through a bottleneck is illustrated in Figure 1.

The models of Brighton and others are called *iterated learning models*. In a nutshell, at each instance in time the population of an iterated learning model contains adults and children. (Typically the models have one adult and one child, but see Smith and Hurford (2003); Vogt (2005a) for studies with larger population sizes.) The adults talk to children, who then learn from the linguistic input they receive (see Fig. 2). After a given number of learning trials (or *language games*), the adults are removed from the population, the

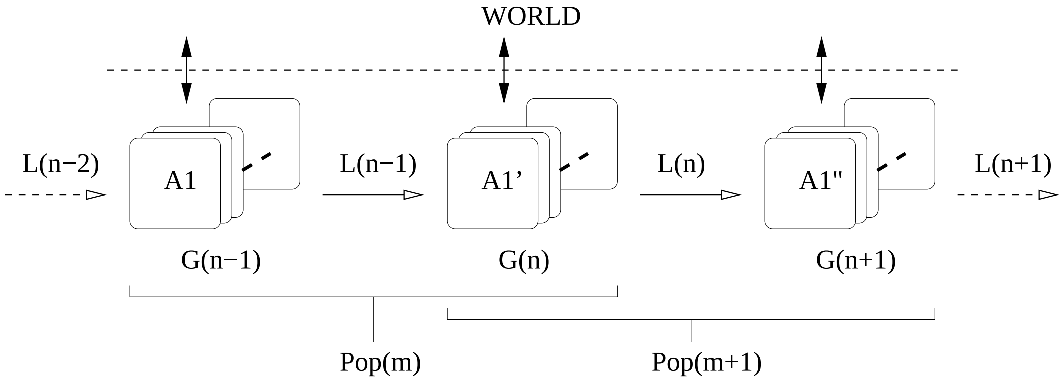

Figure 2: A schematic overview of the iterated learning model. This scheme shows two populations Pop(m) and Pop(m+1), each present at iterations m and m+1. Each population contains agents from two generations, e.g., Pop(m) contains generations G(n-1) and G(n). Each generation, in turn, contains a given number of agents. The agents from the older generation (e.g., G(n-1)) are the adults in the population and the agents from the younger generation (e.g., G(n)) are the children. Each individual agent A has its own private language, which it has acquired from observing the linguistic expressions L from the previous generation. Hence, children acquire the language through linguistic interactions with the adults of that population. In the current model, these linguistic interactions are based on the agents' perception of the world. At a given moment, all adults are removed, the children become new adults and new children enter the population and the cycle repeats.

children become the new adults and new children enter the population; this process repeats indefinitely. What basically happens in these iterated learning models (ILMs) is that the (initially holistic) languages themselves adapt to become learnable for following generations, as suggested by Deacon (1997); Wray (1998). The driving forces of this language change are the transmission bottleneck and the language acquisition mechanisms. In NSL, Senghas et al. (2004) have identified two creative learning mechanisms that they hold accountable for the children's creativity:

1. Breaking apart previously unanalysed wholes.
2. A predisposition for linear sequencing.

These learning mechanisms are also central to the iterated learning models (ILMs) of Brighton (2002); Kirby (2001); Smith et al. (2003); Vogt (2005a), and are similar to those proposed by Lieven et al. (2003); Tomasello (2000); Wray (1998). The next section illustrates such a learning mechanism in more detail.

Many computational models regarding the evolution of language have focused on the emergence of grammatical structures, see, e.g., (Briscoe, 2002; Cangelosi and Parisi, 2002) for overviews. Most of these models take the development of semantics for granted, but recently several models have been developed in which the semantics are grounded from the interaction of the agents with their environment (Steels, 2004; Vogt, 2005a). In the current study, which uses the model introduced in Vogt (2005a), the meanings are also developed ontogenetically from the interaction of agents with their environment. Although not unimportant, the current paper does not focus on the effect of ontological development of meanings on the emergence of compositionality; see (Vogt, 2005a,b) for a discussion. Instead, this paper focuses on the productive creativity of children (a.k.a. *learners*) when they are allowed to speak while they are still in their developmental phase.

As mentioned, the ILM typically assumes a vertical transmission of language. Based on definitions discussed in Cavalli-Sforza and Feldman (1981), the term *vertical transmission* is used to denote a transmission from any individual of one generation to any individual of the next. (Note that in Cavalli-Sforza and Feldman (1981) this type of transmission is coined *oblique transmission*, they use vertical transmission to denote transmission from a parent to its offspring.) In contrast, the term *horizontal transmission* is used to denote transmission between any two (usually unrelated) individuals. (The latter terminology is taken from epidemiology. Note that Cavalli-Sforza and Feldman (1981) takes horizontal transmission between any two individual from the same generation.)

In all ILMs studied so far, only adults talk to children, i.e. the language is transmitted *vertically*. This has the consequence that the experimenter has to impose a bottleneck explicitly in order for compositionality to arise. However, when a child is allowed to speak earlier in life, it may naturally want to communicate about a meaning it has never encountered before. This implicitly imposes a bottleneck for the child, which may give rise to using a compositional structure. For example, when the child has already learnt from observing the expressions “green ball”, “red ball” and “red car”, it is – provided it has the right generalisation mechanisms – capable of producing the utterance “green car” despite it never communicated about green cars before.

The early productivity is important for grammatical development, because it has been argued that in order

> “[t]o become a competent speaker of a natural language it is necessary to be conventional: to use language the way that other people use it. [... However,] it is also necessary to be creative: to formulate novel utterances tailored to the exigencies of particular communicative circumstances.” (Tomasello, 2000, p. 209)

Young children do show the use of expressions they never heard before. One famous example is a child’s utterance “Allgone sticky” reported by Braine (1971) (cited in Tomasello, 2000). It seems that this child has used the form “sticky” as a generalisation for “juice” and “paper”, which the child previously heard in utterances such as “Allgone juice” or “Allgone paper”. This new generalisation was then substituted to form the expression “Allgone sticky”. It is noteworthy that children only tend to produce expressions based on what they have heard before Lieven et al. (2003). This is also true for the current model. Although children in the model are allowed to invent new word-forms, they also tend to substitute new items based on previously chunked rules. As we shall see, such substitutions indeed give rise to the emergence of compositionality in simulations without an explicitly imposed transmission bottleneck, even though this would not occur under the same absence of a bottleneck when there is only vertical transmission.

The study presented in Vogt (2005a), which used the same model as in this paper, also implemented the ILM based on vertical transmission. In that paper, the model, which combines the ILM with Steels’ (1996a) language game model and some new induction mechanisms, revealed that compositionality under all circumstances emerged quite rapidly – even in the absence of a bottleneck. Although this compositionality remained stable over subsequent generations when a bottleneck was imposed explicitly, when no bottleneck was imposed, compositionality tended to die out and was replaced by holistic languages. This result differed from the results presented by, e.g., Brighton (2002); Kirby (2001); Smith et al. (2003), where compositionality usually did not emerge in the first place. The reason for this difference was explained by the statistical nature of the data (both linguistic and perceptual), the differences in learning mechanisms, and the ontological development of meanings (Vogt, 2005a,b). Inherent to the parameter settings of this model, the chance of finding regularities in linguistic signals on one hand and perceptual features on the other was much higher, which gave rise to a rapid development in compositionality in the first place. The different learning mechanism used in Vogt’s model and the asynchronous development of meaning (i.e. children develop meanings from scratch while adults have a fully developed semantics) has shown to give further rise to the rapid development of compositionality. Despite these differences, the model reveals qualitatively similar results. Given these results, it is predicted that when the model has no bottleneck but includes more horizontal communication (i.e. children are allowed to speak as well), compositionality will remain stable over subsequent generations. The remainder of this paper shows how this is achieved in the computer model.

speaker	hearer
perceive context	perceive context
categorise objects	categorise objects
select topic	–
discrimination game	discrimination games
encode expression	decode expression
–	indicate topic
provide feedback	evaluate feedback
adapt weights	adapt weights
–	induce language

Table 1: An outline of the guessing game. See the text for details.

3 Guessing games

The model is based on the Talking Heads experiment (Steels et al., 2002) in which a population of agents could embody themselves into a robot that was set up as a pan-tilt camera connected to a computer. These robots looked at parts of a white board on which geometrical coloured shapes were pasted. The population engaged in a series of language games (Steels, 1996a) (or, more specifically, *guessing games*) to develop a shared lexicon with which they could communicate about the objects on the white board. In the original Talking Heads experiment, there were several setups located in different laboratories across the world which were connected to each other through the Internet. The current study is implemented in a simulation toolkit of the Talking Heads, called *THSim* (Vogt, 2003).[1]

The communication between agents is modelled through guessing games, which is a type of language game based on corrective feedback.[2] In a guessing game, two agents participate: a speaker and a hearer which are selected from the population according to some rule, specified in Section 4. As outlined in Table 1, both agents *perceive a context*, which consists of a number of geometrical coloured objects randomly selected from a given world of objects. For each object, the agents extract a feature vector, which they *categorise* using a mechanism based on the *1-nearest neighbourhood algorithm* (Cover and Hart, 1967). The speaker selects one object as the *topic*, which the hearer has to guess based on the expression that the speaker will produce. In order to form a meaning of the topic, the speaker plays a *discrimination game* (Steels, 1996b). As the hearer does not know which object is the topic, it will play a discrimination game for each object in the context. The speaker then tries to *encode* an expression; if this fails, some new linguistic rule may be *invented*. In turn, the hearer will try to *decode* the expression. If this succeeds, the hearer has guessed the speaker's topic, which is then *indicated* to the speaker. The speaker verifies whether this is the intended topic and provides (corrective) *feedback* to the hearer. The hearer then *evaluates this feedback*. Based on the outcome of the game, the agents then *adapt weights* that are associated with the rules in their grammar. However, if the hearer has failed to decode an expression or if it guessed the wrong topic, it will try to induce new rules based on the speaker's utterance and the topic, which is then provided by the speaker as part of the corrective feedback. This would allow the hearer to encode or decode the expression in future occasions.

In the remainder of this section, the model is explained in some more detail. However, due to the complexity of the model and a lack of space, some explanations will be very brief or not well motivated. For more details on the model and the motivation of design choices, consult (Vogt, 2005a).

[1] THSim is freely downloadable from http://www.ling.ed.ac.uk/~paulv/thsim.html and includes its source code written in Java.

[2] THSim also includes an implementation of the *observational game*, which is based on joint attention and Hebbian learning, but which performs significantly worse than the guessing game, see (Vogt, 2005a). A third game, based on cross-situational statistical learning (Vogt and Smith, 2005), is currently only implemented in THSim to simulate lexicon formation.

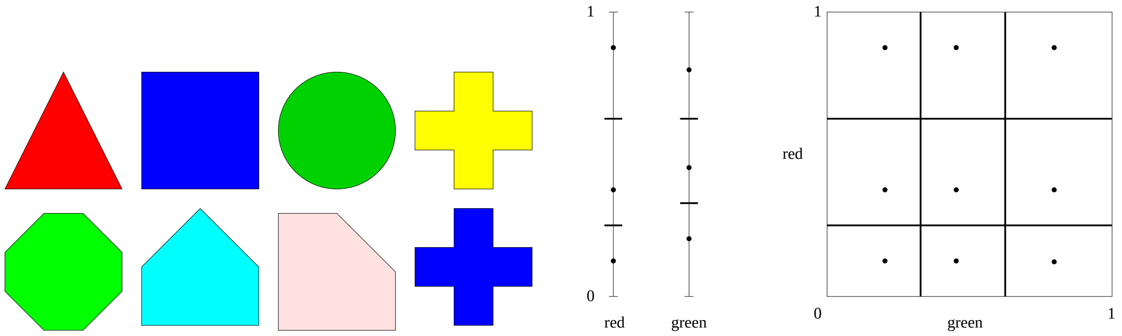

Figure 3: **Left:** An example context for a language game. The context consists of 8 objects formed from combining one of 10 shapes (e.g., a circle, triangle, square, ...) with one of 12 fixed colours (e.g., red, yellow, green, ...). **Right:** This figure shows an example of how an ontology can be interpreted and used in terms of a conceptual space. There are two quality dimensions **red** and **green**, each with three categorical features (points) that form 3 regions in each dimension (left). When combined, these dimensions form a conceptual space that has 9 different categories (right). Note that the number of categorical features is not fixed, nor is the number of different conceptual spaces (i.e. it is not specified which quality dimensions should be combined to form a conceptual space).

3.1 Context perception, categorisation and discrimination

The guessing games are situated in a context that consists of 8 objects randomly selected from the world. The world contains 120 objects made from 12 colours combined with 10 shapes. An example context is shown in Fig. 3. For each object, a 4-dimensional feature vector is obtained that has three perceptual features for colour (the normalised components of the RGB colour space) and one perceptual feature for shape (calculated from the ratio between the object's area and the area of its smallest bounding box).[3]

Once the agents have obtained a feature vector for each object, they will relate each vector to a category. A category is formed from 4 *categorical features*, which the agents store in their *ontologies*. (Note that initially the agents' ontologies are empty.) A categorical feature is a focal point (or *prototype*) in one quality dimension of each agent's conceptual space (cf. Gärdenfors, 2000). Each feature is related to its nearest categorical feature. Combining all 4 thus related categorical features forms the category, see Fig. 3.

When all objects are categorised this way, the individual agent plays a discrimination game for the topic (speaker) or for each object in the context (hearer). The aim of the discrimination game is to verify whether the category of the object under consideration distinguishes this object from the rest of the context. If this is the case, the discrimination game succeeds and the language game proceeds with this *distinctive category DC*. If not, the game fails and the agent adds new categorical features to its ontology for which the topic's features serve as exemplars, insofar the categorical features do not yet exist. If the speaker's discrimination game or all discrimination games played by the hearer fail(s), the guessing game fails as well.

3.2 Encoding and decoding

The categories of the agents are constructed within *conceptual spaces* (Gärdenfors, 2000). A conceptual space is spanned by a number of *quality dimensions*, such as the `r`, `g` and `b` components of the RGB

[3]Note that although agents categorise objects in terms of colour and shape, this study does not investigate the formation of human colour categories as in Belpaeme's studies (Belpaeme and Bleys, 2005; Steels and Belpaeme, 2005). For this, the used representation of colours (i.e. the RGB colour space) is too unrealistic.

R_1	`S -> aba/`$[1_r^{0.1}, 0_r^{0.3}, 0_g^{0.4}, 0_b^{0.1}, 1_b^{0.3}, 1_s^{0.4}]$	0.1
R_2	`S -> A/rgb B/s`	0.8
R_3	`A -> a/`$[1_r^{0.3}, 0_g^{0.3}, 0_b^{0.3}]$	0.3
R_4	`A -> db/`$[0_r^{0.4}, 0_g^{0.4}, 1_b^{0.4}]$	0.4
R_5	`B -> ka/`$[1_s^{0.2}, 0_s^{0.3}]$	0.4

Table 2: An example grammar. In between the square brackets are categorical features listed with their feature value, dimension and category weight. For instance, the categorical feature $1_r^{0.1}$ has a feature value of 1 for dimension r and weight 0.1. The dimensions r, g, b and s stand for red, green, blue and shape respectively. The final column gives the rule weight ρ_i.

colour space (denoted by `rgb`). In this study, a holistic conceptual space is spanned by all 4 dimensions and compositions are formed from spaces spanned by a subset of these dimensions. Sentences are formed from a (composition of) conceptual space(s) that contain all 4 dimensions once.

Each agent constructs during its 'lifetime' a grammar that consists of simple rewrite rules of the form `S -> A/x B/y` for compositional rules or `S -> word/meaning` for holistic rules, such as shown in Table 2. In the compositional rules, `A`, `B`, `C`, ... are non-terminals that *cover* a conceptual space (here denoted by `x` and `y`). The grammar also has non-terminal rules of the form `A -> word/meaning`. The holistic rules and non-terminal rules rewrite to word-forms `word`, which are strings constructed from letters taken from a predefined alphabet, and meanings `meaning`, which are lists of categorical features. In principle a `meaning` can have more than one categorical feature in each dimension which the rule covers. This may give rise to ambiguities or overextensions which can be disambiguated or narrowed down through the adaptation of weights attached to each categorical feature of a rule. In the current implementation, only sentences `S` rewrite to compositional rules, which are limited to two constituents only.

3.2.1 Encoding

When a speaker tries to encode an expression, given a distinctive category DC_t for topic t, it looks for those compositions whose meanings match DC_t. For instance, if $DC_t = (1_r, 0_g, 0_b, 1_s)$, this matches the compositions given by R_1 and $R_2 \circ R_3 \circ R_5$ in Table 2. (Note that the operator $\circ$ is used to bind rules in a composition; $\oplus R_i$ is used to denote a shorthand for the composition.) As a distinctive category can decode into more than one expression, the speaker will have to select one. To this aim, the speaker calculates a composition score $\sigma(\oplus R_i)$ and selects the composition with the highest score. Each rule R_i has a weight ρ_i and each categorical feature c_k associated with a rule has categorical weight $w(c_k)$. To calculate the compositions' scores, the agents first calculates the score s_i for each rule R_i used in the compositions following Eq. (1):

$$s_i = \begin{cases} \rho_i w_i & \text{if } R_i \text{ has specified meanings,} \\ \rho_i & \text{if } R_i \text{ rewrites into non-terminals.} \end{cases} \tag{1}$$

where w_i is the average category weight $w(c_k)$ of the categorical features c_k that match those specified in DC_t. This way, $s_1 = (0.1) \cdot (\frac{0.1+0.4+0.1+0.4}{4}) = 0.025$, $s_2 = 0.8$, $s_3 = (0.3) \cdot (\frac{0.3+0.3+0.3}{3}) = 0.09$ and $s_4 = (0.4) \cdot (0.2) = 0.2$. From these scores, the composition score $\sigma(\oplus R_i)$ of a composition is calculated according to Eq. (2):

$$\sigma(\oplus R_i) = \prod_{\oplus R_i} s_i \tag{2}$$

In the example $\sigma(R_1) = 0.025$ and $\sigma(R_2 \circ R_3 \circ R_5) = 0.8 \cdot 0.09 \cdot 0.2 = 0.0144$. The speaker selects the composition that has the highest score and, in case of a tie, the speaker makes a random choice. So in this example composition R_1 is favoured over $R_2 \circ R_3 \circ R_5$ and the speaker will utter the expression "aba".

If the speaker fails to encode an utterance, this is because it has no rule in its grammar that matches the topic's distinctive category. If this is the case, the speaker invents a new rule, first by searching for a way to *exploit* an already existing rule. If that is not possible, it will create a new holistic rule.

For example, if the speaker wishes to encode an expression for $DC_t = (0_r, 1_g, 0_b, 1_s)$ using the grammar from Table 2, it cannot make a composition that matches this entire category. However, the categorical feature 1_s can be matched with rule R_5, so the speaker can exploit this by adding a new rule covering the RGB colour space with a newly created word to its grammar. New words are created by forming a random string of letters taken from a predefined alphabet with a length l between 2 and 8, where the length l is distributed following Zipf's law, i.e., $f(l) \propto \frac{1}{l}$. The use of Zipf's law implements a bias towards producing short words; a property that has also been observed in human languages (Zipf, 1949). In this example, the added rule (R_6) could be something like `A -> bla/`$[0_r^{w_0}, 1_g^{w_0}, 0_b^{w_0}]$ with initial category weights $w_0 = 0.01$ and rule weight $\rho_0 = 0.01$. This way, the speaker can now produce the expression "blaka" using composition $R_2 \circ R_6 \circ R_5$.

If, for instance, the distinctive category had been $DC_t = (0_r, 1_g, 0_b, 0.5_s)$, no existing knowledge could be exploited and the agent creates a new holistic rule such as `S -> gk/`$[0_r^{w_0}, 1_g^{w_0}, 0_b^{w_0}, 0.5_s^{w_0}]$, with ρ_0. Expressions encoded using newly created words are uttered immediately.

3.2.2 Decoding

When the hearer receives an expression, it will try to decode the expression. To this aim, the hearer searches its grammar for compositions that match the expression and one of the distinctive categories resulting from the discrimination games played on the context's objects. For example, if the hearer receives the utterance "aba", rule R_1 from Table 2 is the candidate rule. Now suppose that the hearer could categorise two objects distinctively into $DC_1 = (1_r, 0_g, 0_b, 1_s)$ and $DC_2 = (0_r, 0_g, 1_b, 1_s)$, then both can be decoded into a composition featuring R_1, because R_1 has ambiguous categorical features in the `r` and `b` dimensions. For decoding DC_1, the composition score becomes $\sigma(R_1)' = 0.025$ (as above) and for decoding DC_2 this becomes $\sigma(R_1)'' = (0.1) \cdot (\frac{0.3+0.4+0.3+0.4}{4}) = 0.035$. The hearer selects the second interpretation, because its composition score is highest.

Now the agents will verify whether or not the hearer guessed the right topic. To this aim, the hearer 'points' at the object relating to the distinctive category belonging to the decoded expression, thus indicating the guessed object to the speaker. If this object is the topic intended by the speaker, the guessing game is successful, if not there is a *mismatch in referent*. This is verified by the speaker and the result of the game is passed back to the hearer. If there is a mismatch in referent or the hearer could not decode the utterance, the speaker additionally provides corrective feedback by indicating its intended object. Although it is disputed whether or not this type of feedback is observed frequently in adult-child communication (Bloom, 2000), it is assumed here that a hearer in such cases is able to find the appropriate reference in one way or the other. Moreover, it has been shown that this strategy of guessing and providing corrective feedback is more effective in forming stable compositional languages than providing the topic prior to the verbal interaction, see Vogt (2005a) for details.

3.3 Adaptation of weights

Upon receiving feedback, the hearer evaluates the success of the game. If successful, the weights of the used rule(s) ρ_i and associated categorical features $w(c_{i,k})$ are increased, while the weights of competing elements are laterally inhibited. These adaptations are carried out by both agents. If there is a mismatch in referent, only the hearer lowers the weights of used elements. Weights are increased following:

$$w = \eta \cdot w + (1 - \eta) \tag{3}$$

where $\eta = 0.9$ is a constant learning parameter and $w = \rho_i$ or $w = w(c_{i,k})$. Lowering (or inhibiting) weights is done using the following equation:

$$w = \eta \cdot w \tag{4}$$

This way, all weights remain bounded in the range $\langle 0, 1 \rangle$. Note that the lateral inhibition of the rule weight ρ_j helps to disambiguate different grammatical structures, while the lateral inhibition of the categorical features' weights $w(c_j, k)$ helps to disambiguate the meaning of a rule. Rules are said to *compete* if they are part of a composition that parsed the distinctive category (speaker) or utterance (hearer), but were not selected. Categorical features compete if they are part of a selected rule, but did not contribute to its meaning.

This way, successfully used rules and meanings are more likely to be reselected in future games, while unsuccessful ones are less likely to be used in the future. It has been shown that that such update mechanisms serve a self-organisation of language, see, e.g., De Jong (1999).

3.4 Induction

At the start of each agent's lifetime, its grammar is empty, so if decoding fails, the agents need to induce new knowledge. When the hearer is unable to decode the speaker's expression or if there is a mismatch in referent, it will expand its grammar by trying to induce new rules. Before doing so, the speaker 'points' at its intended topic, thus providing corrective feedback. In all following cases, the hearer is thus aware of the topic. There are three different mechanisms that are considered in the following order *exploitation*, *chunking* and *incorporation*. The algorithm stops if one of the mechanisms succeeds. After these methods are applied, two post operations *generalise* and *merge* are carried out. These induction mechanisms are largely based on Kirby's models (Kirby, 2001, 2002), though with a few differences not discussed in the current paper, see Vogt (2005a) for details.

Exploitation: If a part of the expression and a part of the topic's distinctive category can be decoded by the hearer's grammar, the hearer can exploit this knowledge to construct a rule that would allow it to decode the entire expression in a future occasion. For instance, if the expression is "toka" and $DC_t = (0_r, 1_g, 0_b, 1_s)$, then the part "ka" meaning $[1_s]$ can be decoded using the grammar of Table 2. In this case, the hearer adds the rule `A -> to/`$[0_r^{w_0}, 1_g^{w_0}, 0_b^{w_0}]$, with $\rho_0 = w_0 = 0.01$ to its grammar. If there are more compositions that can be exploited, the agent selects the composition with the highest score $\sigma(\oplus R_i)$ calculated over the known parts of the composition.

Chunking: Chunking is used when an expression cannot be partially decoded with the given grammar. The aim is to find a way to chunk (or *break up*) the expression and the topic's distinctive category into a new compositional rule. To do this, the agent uses *alignment learning* (van Zaanen, 2000) with respect to an instance base that each agent builds during its lifetime. The instance base contains, for each game, an (unanalysed) expression-meaning pair used for the topic, possibly obtained through corrective feedback.

The hearer searches its instance base for the most effective way to chunk the expression and its meaning. The effectiveness is based on finding alignments that will break up the expression-meaning pair such that if such a rule had existed before, it would have been used more frequently than any other chunk. Consider, for example, the instance-base shown in Table 3. Suppose a hearer receives the expression "ekba" together with $DC_t = (1_r, 1_g, 1_b, 0.5_s)$. The following alignments with the expression can be found: "ad$\underline{\text{a}}$", "mk$\underline{\text{a}}$" and "lpt$\underline{\text{a}}$". Note that alignments in the expressions are only allowed at the beginning or the end, so "m$\underline{\text{kb}}$io" and "e$\underline{\text{kb}}$a" do not align. (This is done for computational reasons.) At the syntactic level we can therefore only chunk "ekba" into "ekb" and "a". At the semantic level this constraint is not used, so instance (1) aligns with the distinctive category in three ways, cf. (a), (b) and (c) in Table 3. Similarly, instances (2) and (3) also align in three ways, cf. (d) – (i). The right part of Table 3 shows all possible alignments. Looking carefully, one can see that the alignment "ekb$\underline{\text{a}}$" with $(1_r, 1_g, 1_b, \underline{0.5_s})$ fits with all

Instance-base				Alignments			
1	ad$\underline{\text{a}}$	$(0_r, 1_g, 0_b, 0.5_s)$	3	a	ekb$\underline{\text{a}}$	$(0_r, \underline{1_g}, 0_b, \underline{0.5_s})$	3
				b	ekb$\underline{\text{a}}$	$(1_r, \underline{1_g}, 1_b, 0.5_s)$	3
				c	ekb$\underline{\text{a}}$	$(1_r, 1_g, 1_b, \underline{0.5_s})$	3
2	mk$\underline{\text{a}}$	$(0_r, 0_g, 1_b, 0.5_s)$	2	d	ekb$\underline{\text{a}}$	$(1_r, 1_g, \underline{1_b, 0.5_s})$	2
				e	ekb$\underline{\text{a}}$	$(1_r, 1_g, \underline{1_b}, 0.5_s)$	2
				f	ekb$\underline{\text{a}}$	$(1_r, 1_g, 1_b, \underline{0.5_s})$	2
3	lpt$\underline{\text{a}}$	$(1_r, 0_g, 0_b, 0.5_s)$	2	g	ekb$\underline{\text{a}}$	$(\underline{1_r}, 1_g, 1_b, \underline{0.5_s})$	2
				h	ekb$\underline{\text{a}}$	$(\underline{1_r}, 1_g, 1_b, 0.5_s)$	2
				i	ekb$\underline{\text{a}}$	$(1_r, 1_g, 1_b, \underline{0.5_s})$	2
4	rtd	$(1_r, 1_g, 0_b, 0_s)$	2		–		
5	mkbio	$(1_r, 1_g, 0_b, 1_s)$	2		–		

Table 3: This table shows on the left part an example list of instances (expression-meaning pairs) together with the frequencies with which they occurred (final column of the left part). The right part shows the alignments that the expression "ekba" with $DC_t = (1_r, 1_g, 1_b, 0.5_s)$ has with the entries of the instance base. The final column gives the frequencies with which the alignment is found in the instance-base. See the text for further details.

three instances, cf. (c), (f) and (i); the other alignments only occur with one of the instances. Summing their occurrence frequencies yields 7, which is higher than the frequencies of the other alignments. Based on the occurrence frequencies of the alignments, the hearer chunks the expression "ekba" into "ekb" and "a" and it chunks $DC_t = (1_r, 1_g, 1_b, 0.5_s)$ into $(1_r, 1_g, 1_b)$ and (0.5_s). This gives the hearer three new rules: `S -> A/rgb B/s`, `A -> ekb/`$[1_r^{w_0}, 1_g^{w_0}, 1_b^{w_0}]$ and `B -> a/`$[0.5_s^{w_0}]$, each with an initial weights $\rho_0 = w_0 = 0.01$.

Holistic incorporation: If neither exploitation or chunking is possible, the hearer adopts the expression with its meaning holistically. For example, if the hearer hears "wateva" with distinctive category $DC_t = (0.5_r, 0.5_g, 0.5_b, 0.6_s)$ and assuming Table 2 is the existing grammar, then, whatever the instance base looks like, no rule can be exploited, nor can a chunk be made and the rule `S -> wateva/`$[0.5_r^{w_0}, 0.5_g^{w_0}, 0.5_b^{w_0}, 0.6_s^{w_0}]$ will be added to the grammar with rule weight ρ_0.

Generalise and merge: Two post operations are required to generalise the newly constructed knowledge further and to remove any constructed redundancies. First, if there are already some holistic rules that can make a similar chunk, the grammar is expanded by these generalised chunks. E.g., when the grammar contains rules such as `S -> ada/`$[0_r^{0.2}, 1_g^{0.2}, 0_b^{0.2}, 0.5_s^{0.2}]$ and `S -> ekbor/`$[1_r^{0.3}, 1_g^{0.3}, 1_b^{0.3}, 0.1_s^{0.3}]$ and when through chunking the rules `S -> A/rgb B/s`, `A -> ekb/`$[1_r^{w_0}, 1_g^{w_0}, 1_b^{w_0}]$ and `B -> a/`$[0.5_s^{w_0}]$ were introduced, the rules `A -> ad/`$[0_r^{0.2}, 1_g^{0.2}, 0_b^{0.2}]$ and `B -> or/`$[0.1_s^{0.3}]$ are added to the grammar while keeping the already existing rules.

Second, if there are rules which are in essence the same, these are merged. For instance, if the grammar already contained rules `S -> E/rgb F/s` and `F -> a/`$[0_s^{0.3}]$ when the "ekba" chunk was applied, the rule `S -> A/rgb B/s` is removed and the non-terminal letters `E` and `F` are substituted for `A` and `B` in the non-terminal rules. In addition, the rule (now) `F -> a/`$[0.5_s^{w_0}]$ is merged with `F -> a/`$[0_s^{0.3}]$ to become `F -> a/`$[0_s^{0.3}, 0.5_s^{w_0}]$.

4 Experiments

In this section, one experiment is reported in which the rates with which speakers and hearers are selected from the adult population were varied. In all simulations of this experiment, the population size was set to 6 agents: 3 adults and 3 children. In the experiment, 2 parameters were varied: (1) the probability pS with which a random *speaker* is selected from the adult generation, and (2) the probability pH with

$pS\backslash pH$	0.00	0.25	0.50	0.75	1.00
0.00	0.58 ± 0.04	0.65 ± 0.03	0.65 ± 0.03	0.65 ± 0.02	0.48 ± 0.01
0.25	0.67 ± 0.04	$\mathbf{0.68 \pm 0.02}$	$\mathbf{0.68 \pm 0.04}$	$\mathbf{0.67 \pm 0.02}$	0.12 ± 0.22
0.50	0.72 ± 0.10	$\mathbf{0.58 \pm 0.18}$	$\mathbf{0.57 \pm 0.22}$	$\mathbf{0.66 \pm 0.09}$	0.59 ± 0.01
0.75	0.28 ± 0.14	$\mathbf{0.24 \pm 0.24}$	$\mathbf{0.18 \pm 0.13}$	$\mathbf{0.26 \pm 0.21}$	0.57 ± 0.02
1.00	*0.33 ± 0.14*	0.17 ± 0.09	0.41 ± 0.31	0.48 ± 0.24	0.56 ± 0.05

(a) Compositionality.

$pS\backslash pH$	0.00	0.25	0.50	0.75	1.00
0.00	0.34 ± 0.01	0.55 ± 0.02	0.57 ± 0.03	0.57 ± 0.01	0.26 ± 0.00
0.25	0.61 ± 0.03	$\mathbf{0.72 \pm 0.03}$	$\mathbf{0.70 \pm 0.05}$	$\mathbf{0.66 \pm 0.04}$	0.24 ± 0.12
0.50	0.75 ± 0.07	$\mathbf{0.68 \pm 0.15}$	$\mathbf{0.67 \pm 0.14}$	$\mathbf{0.71 \pm 0.07}$	0.41 ± 0.01
0.75	0.61 ± 0.11	$\mathbf{0.57 \pm 0.13}$	$\mathbf{0.57 \pm 0.06}$	$\mathbf{0.61 \pm 0.10}$	0.42 ± 0.01
1.00	*0.67 ± 0.05*	0.63 ± 0.06	0.73 ± 0.11	0.73 ± 0.11	0.17 ± 0.00

(b) Communicative accuracy.

Table 4: The top table provides the average level of compositionality reached at the end of the simulations plus their standard deviations. Each row contains the results of a different value of pS as specified in the first column, and each column gives the results for a different value of pH as specified in the first row. The values displayed in boldface relate to horizontal transmission; the values in italics to vertical transmission. The bottom table provides the average level of communicative accuracy reached at the end of the simulations together with their standard deviations.

which a random *hearer* is selected from the adult generation. Both parameters were varied from 0 to 1 with intermediate steps of 0.25. If a speaker or hearer is not selected from the adult generation, it is selected from the child generation. Naturally, the hearer cannot be the same agent as the speaker. I refer to the condition where $pS = 1$ and $pH = 0$ as vertical transmission, while horizontal transmission is used to denote cases where $0 < pS < 1$ and $0 < pH < 1$. All other extreme cases involve either partial vertical transmission ($pS > 0$ and $0 < pH < 1$), partial inverse vertical transmission ($pS = 0$ and $0 < pH < 1$, or $pH = 1$), or no transmission ($pS = pH = 0$ or $pS = pH = 1$). Although these extreme cases, apart from vertical transmission, are less interesting for the purpose of this paper, their results are included for completeness.

In all simulations, the population played 6,000 guessing games per iteration as 'training' and were then tested for 200 different situations before the adults were removed from the population, the children became adults and new children were introduced. In each situation of the testing phase, a context was constructed and one object was selected randomly as a topic. Then each agent tried to encode an expression, which in turn each other agent tried to decode. During the testing phase, there was no further learning or invention going on – i.e. both ontologies and grammars remained fixed throughout the testing phase. From the testing phase, two measures were calculated:

Compositionality measures the number of expressions that are encoded or decoded using a compositional rule, normalised by the total number of expressions that were encoded and decoded.

(Communicative) accuracy measures for each agent in each situation how many agents successfully understand this agent's expression, normalised by the number of agents and the number of situations.

Each simulation was repeated 10 times with different random seeds for statistical purposes.

4.1 Results

Table 4 shows the results from simulations in which pS and pH were varied from 0 to 1 with intermediate steps of 0.25. Table 4 (a) shows the average compositionality with standard deviation measured in the test phase of the final iteration. The average communicative accuracies in the test phase of the final iterations in these simulations are given in Table 4 (b).

From Table 4 (a), it is clear that when the majority of speakers are adults (i.e. $pS \geq 0.75$, bottom 2 rows), compositionality is clearly much lower than when children form the majority of speakers (i.e. $pS \leq 0.50$), apart from the case where $pH = 1.00$ (final column). When $pS = 0.50$ and $pH = 0.00$, compositionality is highest. However, the differences between the different values of compositionality when $pS \leq 0.50$ are not significant according to the Wilcoxon rank sum test ($p > 0.1$). The differences between the higher levels of compositionality when $pS \leq 0.50$ (compositionality > 0.55) and the lower values when $pS \geq 0.75$ (compositionality < 0.40) are significant ($p < 0.015$). Whereas when $pS = 1.00$ and $pH \geq 0.50$, the differences with the cases where $pS \leq 0.50$ have lower significance ($0.07 < p < 0.25$).

Note that the case where $pS = 0.25$ and $pH = 1.00$ forms an exception. It is yet unclear what causes this low level of compositionality. However, it should be realised that the extreme values for pS and pH are not very realistic, as in these cases either all speakers are adults or children and/or all hearers are adults or children. In these cases, there are some conditions where the results are completely different than under other conditions.

It should be noted that in the cases where $pS \geq 0.50$, the standard deviations are relatively high. In these cases compositionality did remain stable at a high level in some trials, though typically compositionality evolved to a low level (see, e.g., Fig. 4 (a)).

Table 4 (b) shows that – apart from the cases where $pS = pH = 0.00$ and $pH = 1.00$ – communicative accuracy is relatively high. So, irrespective of whether or not compositional languages remain stable, communication remains effective. This is in line with findings reported by (Smith, 2005, , this issue), who shows how a holistic language can change substantially without affecting its communicative effectiveness.

In short, this experiment shows that, in general, when the majority of speakers are children, compositionality remains stable in the population, whereas when the majority of speakers are adults, compositionality tends to be unstable. Hence, the more vertical the transmission of language, the less likely compositional structures in languages tend to remain if no transmission bottleneck is imposed.

In order to illustrate the difference between vertical and horizontal transmission of language, let us look more closely at what happens in two typical simulations under two different conditions. In the first condition (Fig. 4), the language is transmitted vertically, i.e. $pS = 1.00$ and $pH = 0.00$. In the second condition (Fig. 5), the majority of speakers are children, while the majority of hearers are adults ($pS = 0.25$ and $pH = 0.75$).

Figure 4 shows the results of one typical run with the vertical transmission protocol. As can be seen in the left graph, compositionality developed rapidly in a few iterations to a value of around 0.7, after which it slowly decreased to a value of around 0.4 around the 30st iteration. Around the 60st iteration, compositionality then completely collapsed within a few iterations to a level of around 0.05. Although different runs did not always collapse this way and compositionality sometimes only died out after 200 iterations or so, this type of evolution is typical for vertical transmission schemes when no bottleneck is imposed (Vogt, 2005a). However, when a bottleneck is imposed externally, compositionality tends to increase and remain at a stable level above 0.95, see Vogt (2005a).

The right figure shows for each iteration the relative frequency with which the agents invented a new word as speaker through exploitation (*exploitation(S)*) or holistic invention (*holistic(S)*), or induced new linguistic knowledge as hearer through *exploitation(H)*, *chunking(H)* or holistic *incorporation(H)*. For each of these measures, the number of times a learning mechanism was used was divided by the total number of games within each iteration. This figure clearly shows that the proportion of adaptations carried out

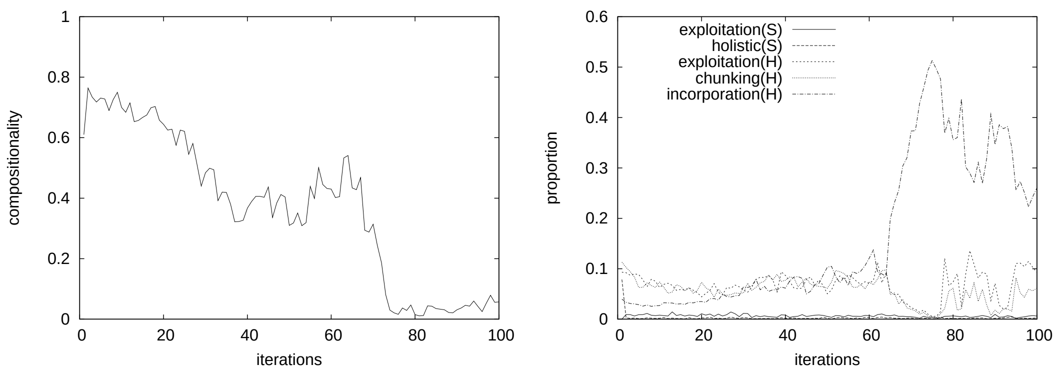

Figure 4: The left figure shows the evolution of compositionality for the vertical transmission case (i.e. $pS = 1$ and $pH = 0$). The y-axis shows compositionality and the x-axis the iterations. The right figure shows the proportion of games per iteration in which the agents invented a new word as speaker S through exploitation (*exploitation(S)*) or holistically (*holistic(S)*), or induced new linguistic knowledge as hearer H through exploitation (*exploitation(H)*, chunking (*chunking(H)* or holistic incorporation *incorporation(H)*.

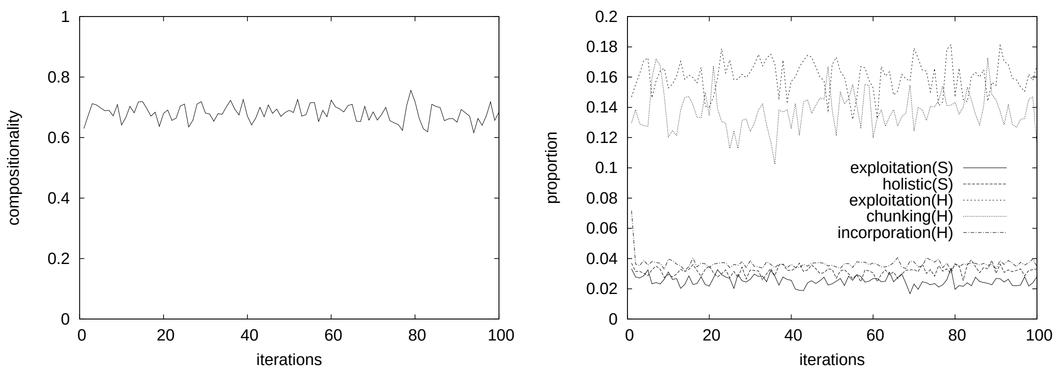

Figure 5: The results of a horizontal transmission case with $pS = 0.25$ and $pH = 0.75$. The left figure shows compositionality. The right figure shows the relative frequency with which different learning mechanisms were applied.

by the hearers – i.e. the children in this condition – (exploiting(H), chunking(H) and incorporation(H)) exceed the proportions of adaptation performed by the speakers (bottom two lines). However, around the moment of collapse, the proportion of adaptations that contribute to the development of compositionality (exploitation and chunking) reduced significantly while the number of incorporation events increased substantially. Apart from the first iteration, where the speakers invented the language holistically, few exploitations and holistic inventions were done by speakers. This is understandable, since the speakers are all adults and have acquired most of the language when they act as speakers.

Figure 5 shows the results of one typical run of a horizontal transmission case. In this simulation, the probability of selecting a speaker from the adult population was $pS = 0.25$ and the probability of selecting a hearer from the adult population was $pH = 0.75$. Compositionality (left) again rose rapidly to a value of around 0.65, but this time remained stable. This tendency was observed in all 10 different trials, cf. the low standard deviation in Table 4.

The relative frequencies of exploitation by the hearers (*exploitation(H)*) and chunking (right figure) were much higher than all other learning mechanisms throughout the evolution. Exploitation by the speakers, holistic invention and incorporation all had similar frequencies between 0.025 and 0.037 on average (see Table 5). Table 5 also shows that all learning mechanisms that give rise to the development

	vertical	horizontal
exploitation (S)	0.006 ± 0.002	0.025 ± 0.003
holistic (S)	0.003 ± 0.008	0.032 ± 0.003
exploitation (H)	0.067 ± 0.027	0.160 ± 0.011
chunking (H)	0.056 ± 0.025	0.138 ± 0.013
holistic incorporation (H)	0.160 ± 0.147	0.037 ± 0.004

Table 5: The relative frequencies with which the different learning mechanisms were used by speakers (S) and hearers (H) for the vertical condition and the horizontal condition. The values presented are averages over the 100 iterations per simulation and their standard deviations.

	vertical	horizontal
substitution (S)	0.155 ± 0.076	0.263 ± 0.053
exploitation (S)	0.116 ± 0.111	0.156 ± 0.053
substitution (H)	0.071 ± 0.056	0.027 ± 0.013
exploitation (H)	0.249 ± 0.078	0.242 ± 0.041
chunking (H)	0.410 ± 0.090	0.312 ± 0.047

Table 6: The proportion of different mechanisms used to encode, decode and learn novel compositional expressions for the vertical and horizontal conditions. The values given are averages over all agents with their standard deviations. The values given for the vertical condition are given for entire lifetime of agents, and the values given for the horizontal condition are given only for children. Note that this way, the substitution (S) and exploitation (S) for speakers in the vertical case are from the agents' adulthood, and the hearer's mechanisms in the vertical case are from the agents' childhood. Further note that since this table only analyses novel compositional expressions, holistic creation and incorporation do not appear.

of compositionality (exploitation and chunking) were used substantially more often in the horizontal condition than in the vertical condition. For instance, the speakers used exploitation roughly 5 times more frequently in the horizontal case than in the vertical case. It must be noted, though, that due to the higher rate of invention (both exploitation(S) and holistic(S))), the languages change more quickly in the horizontal case.

Table 6 shows the proportion of mechanisms used by the agents when encoding, decoding and learning novel compositional structures. In addition to the afore presented learning mechanisms, novel expressions can be formed by *substituting* parts of previously learnt expressions in a compositional rule. Thus such new expressions are used based on recombining previously learnt words. All values give the frequency with which a mechanism is used divided by the total number of new compositional expressions during a given period of an agent's lifetime, averaged over all agents from the two simulations presented above. The designated period for the vertical condition is the agents' entire lifetime; the period measured for the horizontal condition is the agents' childhood period. This way the substitutions and exploitations of the speakers (S) in the vertical condition relates to its adulthood, while their substitutions, exploitations and chunkings as a hearer (H) are from their childhood period.

The table clearly shows that children in the horizontal case frequently use substitutions to form novel expressions, i.e. they recombine previously acquired compositional rules and their slots to form new sentences. Although the mechanisms used as a hearer are lower for the agents of the horizontal condition, it must be noted that during their childhood, the agents of the vertical condition were selected hearer about 4 times as often as the children in the horizontal condition. Furthermore, since the vertical case was measured over agents' whole lifetimes, the figures presented are somewhat skewed. As a consequence, it is likely that, relatively speaking, the hearers of the horizontal case have used substitutions at least as frequent as the hearers of the vertical case. The same holds for chunking. This, however, has not been verified with the data.

Table 7 shows a small fragment of the guessing games played by agent a22 during its childhood. In particular, the fragment shows first time uses of expression-meaning pairs. All games are from the fourth

gg	speaker			hearer		
	A	expr	meaning	A	expr	meaning
28	a22	kje	$[1.0_r, 0.7_g, 0.7_b, 0.0_s]^3$	a18	k,je	$[1.0_r, 0.7_g, 0.7_b][0.0_s]^3$
188	a19	iegmgab	$[1.0_r, 0.0_g, 1.0_b, 0.1_s]^1$	a22	iegmgab	$[1.0_r, 0.0_g, 1.0_b, 0.1_s]^4$
312	a20	i,fmee	$[1.0_r][0.8_g, 0.0_b, 0.5_s]^1$	a22	i,fmee	$[1.0_r][0.8_g, 0.0_b, 0.5_s]^3$
330	a22	i,bonoki	$[1.0_r][1.0_g, 0.0_b, 0.8_s]^2$	a20	i,bonoki	$[1.0_r][1.0_g, 0.0_b, 0.8_s]^2$
535	a23	feflf	$[0.0_r, 1.0_g, 1.0_b, 0.7_s]^1$	a22	f,eflf	$[0.0_r][1.0_g, 1.0_b, 0.6_s]^3$
611	a23	gli	$[0.0_r, 0.0_g, 0.0_b, 1.0_s]^1$	a22	g,li	$[0.0_r][0.0_g, 0.0_b, 1.0_s]^3$
658	a22	g,egmgab	$[0.0_r][0.0_g, 1.0_b, 0.1_s]^1$	a23	g,egmgab	$[0.0_r][0.0_g, 1.0_b, 0.1_s]^3$
661	a22	g,bonoki	$[0.0_r][1.0_g, 0.0_b, 0.8_s]^1$	a20	g,bonoki	$[1.0_b][0.0_r, 0.0_b, 0.8_s]^2$
822	a21	fje	$[0.5_r, 0.5_g, 0.5_b, 0.9_s]^3$	a22	f,je	$[0.3_r][0.7_g, 0.7_b, 0.9_s]^3$
939	a23	i,old	$[0.0_b, 0.0_s][1.0_r, 1.0_g]^1$	a22	i,old	$[1.0_r][1.0_g, 0.0_b, 0.0_s]^2$
1454	a22	f,old	$[0.0_r][1.0_g, 0.0_b, 0.0_s]^1$	a19	f,old	$[0.0_r][1.0_g, 0.0_b, 0.0_s]^1$

Table 7: This table shows a number of guessing played by agent a22 during its childhood in the horizontal condition. The first column indicates the game number, the second to fourth column provide the speaker information, and columns five to seven provide the hearer information. For each agent, the first column gives the agent's identity, the second its expression (if compositional, the two constituents are separated by a comma), the third its meaning (parts). The numbers in superscript indicate which mechanism is used. For the speaker: 1 indicates successful encoded, 2 exploitation and 3 holistic creation. For the hearer: 1 indicates successful decoded, 2 exploitation, 3 chunking and 4 holistic incorporation.

iteration of the horizontal simulation discussed above. The first column of this table indicates the game number within this iteration; the left part of the table shows the agent's identity, expression and meaning for the speaker; and the right part shows the same information for the hearer. Compositional utterances are represented with a comma between the parts of the expression and the corresponding meanings between square brackets in the same order. (Note that agents do not observe these word boundaries.) So, for instance in game 28, agent a18, who is an adult, has interpreted the expression "kje" of the speaker in parts "k,je", where "k" has meaning $[1.0_r, 0.7_g, 0.7_b]$ and "je" has meaning $[0.0_s]$ (the superscript 3 indicates that the hearer has interpret the expression using the chunking mechanism). In this game, agent a22 has encoded the expression using holistic creation. As mentioned, creating new words in the horizontal simulation is quite common, since many speakers are children and they may not have acquired an association between a meaning and an expression yet. In the current experiments, a new word is invented every time this happens. Naturally, this gives rise to a high level of language change. It is possible to assign a probability that regulates the rate with which new words are invented. Preliminary results of experiments investigating such a probability, indeed, reveal less language change and, moreover, a slight increase in compositionality.

All compositional rules of a22 in this example are of the form `S -> A/r B/gbs`, i.e. a composition combining the red component of the colour space with the conceptual space made up of the green and blue components, and the shape dimension. Although one might expect to find most compositional rules of the form `S -> C/rgb D/s` (i.e. colour and shape), statistically the likelihood of finding a regularity in the `r` space is higher than finding one in the `rgb` space (Vogt, 2005b). Nevertheless, on average there is a tendency that rules which break up the colour space and the shape space are most common (Vogt, 2005a).

Game 312 shows an instance of chunking by agent a22. In this game, the hearer heard the expression "ifmee" (note that the hearer did not receive a word-boundary). However, since in game 188, a22 holistically incorporated the expression "iegmgab" meaning $[1.0_r, 0.0_g, 1.0_b, 0.1_s]$ from adult agent a19, agent a22 could find the alignment "ifmee" and $[1.0_r, 1.0_g, 0.0_b, 0.8_s]$, thus breaking up the expression as indicated. While doing this, a22 also chunked the expression "iegmgab" into "i" and "egmgab" with the corresponding meanings. The form "i" is reused in two other games shown: in game 330 a22 exploited the rule using "i" and inventing the form "bonoki"; in game 939, the same rule with expression "i" is exploited to adopt the word "old" from agent a23 – another child in this generation.

Game 611 shows the entrance of the words “g” and “li” in a22’s vocabulary, again a22 used chunking to break up the holistic utterance produced by a23. In games 658 and 661, we see how a22 used substitution to form the expressions “g,egmgab” and “g,bonoki”. A similar process is shown games 822, 939 and 1454. In game 822, our agent acquires the words “f” and “je” through chunking; in game 939 it acquires the word “old” by exploiting the rule with word “i”. Note that in this game, the speaker a23 used a rule of the form `S -> E/bs F/rg`, whereas a22 used rule `S -> A/r B/gbs`. Such misinterpretations do occur occasionally and are based on the differences in the history of language games. Further note that here the word “je” has a category in the conceptual space `gbs`, whereas in game 28 the same word was associated with the shape space `s`.

In game 1454, a22 for the first time produced the expression “f,old” by substituting the previously acquired words “f” and “old”. Surprisingly, a19 – who is an adult – successfully interpreted the expression using the same rule and same constituents, even though it was the first time a19 heard this expression! (Note that this was observed when inspecting the data more closely; the example does not show this.) Agent a19 must therefore have used substitution to decode the expression successfully.

The above examples illustrate most of the properties in the model concerning encoding, decoding, invention and induction mechanisms. Although the exact working of the model is more complicated, the examples should give a feel of how it works.

5 Discussion

In this paper the productive creativity of children with respect to the emergence and evolution of compositionality in language is investigated. To this aim a comparison is made between a vertical model of language transmission and a number of horizontal models of language transmission. It was hypothesised in Section 2 that the horizontal models are better capable at arriving at stable compositional languages than the vertical model when no explicit bottleneck is imposed on the transmission of language. It was argued that this is because in the horizontal models, children start to speak at a stage where they have not yet observed the entire language. So, they need to produce expressions when their input has been sparse, i.e. there is an *implicit bottleneck* on the transmission. As a consequence, children may need to speak about previously unseen objects, which can be done effectively by exploiting regularities they find in the expression and meaning spaces. In turn, this gives rise to the formation of compositionality in their languages (provided, of course, they have the proper learning mechanisms).

This hypothesised effect is, indeed, observed in the experimental results. When the majority of speakers are children, compositionality remains stable throughout the evolution, whereas when the majority of speakers are adults, this is less so. This is an important result, because previous studies using the iterated learning model (see, e.g., Brighton, 2002; Kirby, 2001; Smith et al., 2003; Vogt, 2005a) did not show this aspect, as these models only used vertical transmission protocols and therefore require a bottleneck imposed by the experimenter. The horizontal model is more natural, since it is well known that children start to produce utterances early in their lives and there is typically no external control that sets the subset of the language a child observes. Although it is fair to assume that children never observe all possible utterances of their language, they do seem to observe all possible types of constructions in their language during development (cf. Lieven et al., 2003; Tomasello, 2000).

It is possible to estimate the size of the implicit bottleneck at each new language game a child plays using the *coverage* measure proposed by Brighton (2002). Assuming in each game one object is selected at random as the topic, coverage c estimates the proportion of objects seen by an agent after this played R language games following

$$c = 1 - (1 - \frac{1}{N})^R, \tag{5}$$

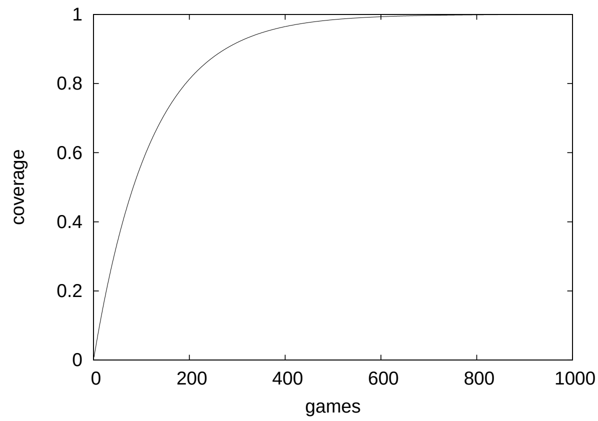

Figure 6: Coverage as a function of the number of language games. See the text for details.

where N is the total number of objects. Figure 6 shows the coverage as a function of the number of language games played. The figure shows that, in the simulations presented above, it takes around 1,000 language games before an agent is likely to have communicated – either as a speaker or hearer – about all 120 objects in the world. When the agent has played around 200 games, it has covered approximately 80% of the objects. So, clearly, the children in the simulations had to communicate about previously uncovered objects during the 1,000 games it plays, thus showing the implicit bottleneck.

Due to the adaptation of scores and the resulting competition among rules, the more a rule is used successfully, the higher its weight becomes. Since the compositional rules are more generally applicable, they are more frequently used, which increases the likelihood to be selected again in favour of a holistic rule that may encode or decode for the same meaning. In addition, the more often compositional rules are used, the more likely other agents will discover such regularities, which again gives further rise to the construction of compositional rules. This competitive selection thus serves a self-organising effect on the formation of compositionality in a similar way ant paths are formed (Prigogine and Strengers, 1984).

Now, how do the results presented in this paper fit with the findings concerning the emergence of Nicaraguan Sign Language? Before trying to answer this question, it is important to place a number of critical remarks about shortcomings of the current model with respect to the situation of NSL. First, the model is no more than a qualitative model of the emergence of NSL. This is necessary, because the state-of-the-art in modelling language evolution is far from realistic with respect to human language evolution. Furthermore, the model was not primarily designed to investigate the emergence of NSL, but since there appear some similarities with NSL, a qualitative comparison is appropriate. So, the model is a very crude abstraction from the human case. For instance, the signals are not gestures, but just some random strings. In addition, the complexity of the signals and their possible meanings are relatively low, because the agents can only communicate about coloured shapes.

As a result of this low complexity, the agents rapidly achieve to construct compositional structures in their languages. This can be explained by realising that statistically there appear many combinatorial regularities in both the signal space and the meaning space, which – given the induction mechanisms – can rapidly be generalised to form compositional rules (see Vogt, 2005a,b, for a detailed explanation). Children acquiring NSL – like all humans – communicate about all sorts of things and their potential space of meanings is infinite. Current ongoing work focuses on developing a model that can account for more complexity in the language by adding more perceptual features, actions and the ability to form more complex sentences.

Another result of this low complexity is that the languages do not become more complex over subsequent generations, which is the case for NSL (Senghas et al., 2004). Instead, in the horizontal case compositionality already emerges in the first generation after which the language stabilises. Apart from the low complexity, one reason may be that there is no natural flow in the population of language users. In the NSL community there is a yearly influx of children who, of course, live longer than two years. The current model only simulates two years of a NSL user. Future work should investigate what happens if there is a more natural influx of children in the model that allows for more complexity in the language.

In the NSL situation, the first group of deaf people brought together already had some communication system in place. Although these homesigns were typically not shared, these people did have more communication skills than the first generation in the horizontal simulations, where the agents developed a shared language from scratch. In the vertical condition, the three adults of the first generation independently invented their own language, which became shared only in later generations. These languages could be considered as homesigns, so in a more realistic scenario, the horizontal condition could be preceded by one generation in which the first generation construct their own private language.

It is also important to note that the population size of 6 agents in the current study is unrealistically low. A more recent study has shown that when increasing the population size to 100 agents in simulations where the speakers and hearers were equally likely to be selected from the adult and child populations (i.e. $pS = pH = 0.5$), compositionality increased to levels around 0.95 and accuracy to levels around 0.9 in 6 out of 10 different runs (Vogt, 2005c). At the end of the 100st iteration of over 200,000 language games each, the average compositionality over 10 runs was 0.76 ± 0.36 and accuracy was 0.70 ± 0.29. In addition, simulations revealed that when the population size is varied between 10 and 50, the results drop significantly with respect to the results presented in this paper. Hence it seems that a large population adds a substantial, but positive pressure to construct compositional languages. It is not yet fully understood why this is the case, but there are many indications that a lot of variety in the private languages puts a pressure on the formation of a well structured language. This is – in a way – related to Darwinian selection, which also depends heavily on variation (Darwin, 1968). Having a lot of variation increases the likelihood of finding a good structure, which can be selected to form stable and coherent languages. Moreover, such coherent languages must be selected in order to become successful in communication.

In the experiments presented in this paper, the rates with which speakers and hearers are selected from the adult and child populations were varied structurally. As the results show, there is some difference in results for the different probabilities. It would therefore be interesting to know what the typical values for such parameters are found in human societies. However, I am unaware of any study that provides such data.

Despite the abstractions and other differences of the current model with respect to NSL, the results do indicate that the productive creativity of children can give rise to the emergence of compositional languages as hypothesised by Senghas and Coppola (2001); Senghas et al. (2004). The increase in the number of exploitations and the high number of new substitutions used by speakers in the horizontal experiments show that when children are the majority of speakers, the level of compositionality increases and remains stable over time. The results are also in line with the findings of Lieven et al. (2003), who have shown that children are very creative with using novel expression based on recombining (or substituting) previously heard utterances.

So the horizontal model increases the creativity among children. However, this creativity also increases the rate with which new words are invented, because each time a meaning is unknown a new word is invented. Consequently, the language changes very quickly, which potentially decreases the accuracy. Nevertheless, such a decrease is not observed. Clearly, the model is strong enough to allow a lot of language change while maintaining a high degree of communicative accuracy. This effect has also been observed in simulations on lexicon formation (Smith, 2005, this issue). As mentioned, one way of dealing with this high level of language change is to reduce the probability with which children create new words when they do not have a lexical entry.

6 Conclusions

The simulations in this paper show that when children are allowed to be productive early in their lifetimes, they are better capable at developing compositional structures in language than when they are more inclined to speak when they become adults, at least when there is no bottleneck imposed by the experimenter. This result is in line with findings from research on children's language acquisition (Lieven et al., 2003; Tomasello, 2000) and the development of Nicaraguan Sign Language (Senghas et al., 2004). The results also support the idea that children are the driving force in the formation of grammar in Creoles, as suggested by Bickerton (1984); Sankoff and Laberge (1973).

The simulations further show that compositionality in language can arise without the need to impose an explicit bottleneck on the transmission as shown by Brighton (2002); Kirby (2001); Smith et al. (2003); Vogt (2005a). Nevertheless, the explanation of transmission bottlenecks as a driving force for the emergence and stability of compositionality as proposed by Kirby (2001) still holds. This is because the children in the simulations presented here face an implicit bottleneck when they start to speak during development.

Like the results presented by Brighton, Kirby and many others, the model supports the idea of, e.g., Deacon (1997); Wray (1998) that languages adapt to the needs of learners, rather than that learners adapt to learn the language as proposed by nativists such as Pinker and Bloom (1990). Moreover, although no sparsity of input is explicitly imposed, the importance of sparsity in the development of compositionality does not necessarily require a nativist stance, as was argued by Chomsky (1980) based on his poverty of the stimulus argument.

The current line work is not yet finished. For instance, in order to make a more reliable comparison to the emergence of NSL, it is important that a similar study is set up with conditions that more closely match those observed in NSL. In addition, work is underway to extend the current model such that more complex languages can emerge, while the recently started New Ties project[4] aims at applying the current model in a large scale simulation that investigates how a potentially large population of artificial agents can evolve a cultural society – including language – based on biological evolution, individual learning and cultural evolution (Gilbert et al., 2005; Vogt and Divina, 2005).

At any rate, if the findings in this study prove to be generally applicable, this would suggest that children may have been the primary driving force in the emergence and evolution of grammatical structures in language. As such, the current study is – I believe – an important step forward in our understanding of the evolution and acquisition of language.

Acknowledgements

This research is funded by a Marie Curie grant awarded by the European Union and a VENI project awarded by the Netherlands Organisation for Scientific Research (NWO). The author is grateful for the invaluable comments received from members of the LEC – in particular Andrew D.M. Smith and Kenny Smith – and two anonymous reviewers.

References

Belpaeme, T. and Bleys, J. (2005). Explaining universal colour categories through a constrained acquisition process. *Adaptive Behavior*. In press.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences*, 7:173–212.

[4]http://www.new-ties.org.

Bickerton, D. (1990). *Language and Species.* Unversity of Chicago Press, Chicago.

Bloom, P. (2000). *How Children Learn the Meanings of Words.* The MIT Press, Cambridge, MA. and London, UK.

Braine, M. D. S. (1971). The acquisition of language in infant and child. In Reed, C., editor, *The Learning of Language.* Scribners, New York.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54.

Briscoe, E. J., editor (2002). *Linguistic evolution through language acquisition: formal and computational models.* Cambridge University Press, Cambridge.

Cangelosi, A. and Parisi, D., editors (2002). *Simulating the Evolution of Language.* Springer, London.

Cavalli-Sforza, L. L. and Feldman, M. W. (1981). *Cultural Transmission and Evolution: A quantitative approach.* Princeton University Press, Princeton, NJ.

Chomsky, N. (1980). Rules and representations. *The behavioral and brain sciences*, 3:1–61.

Cover, T. M. and Hart, P. E. (1967). Nearest neighbour pattern classification. *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, 13:21–27.

Darwin, C. (1968). *The Origin of Species.* Penguin Books.

De Jong, E. D. (1999). Analyzing the evolution of communication from a dynamical systems perspective. In Floreano, D., Nicoud, J.-D., and Mondada, F., editors, *Proceedings of the Fifth European Conference on Artificial Life ECAL'99*, Berlin. Springer-Verlag.

Deacon, T. (1997). *The Symbolic Species.* W. Norton and Co., New York, NY.

Frege, G. (1892). On concept and object. Reprinted in B. McGuiness, ed. Collected Papers on Mathematics, Logic, and Philosophy. Oxford: Blackwell, 1984: 182–194.

Gärdenfors, P. (2000). *Conceptual Spaces.* Bradford Books, MIT Press.

Gilbert, N., Schuster, S., den Besten, M., and Yang, L. (2005). Environment design for emerging artificial societies. In *Proceedings of AISB 2005: Socially inspired computing joint symposium.* In press.

Goldin-Meadow, S. (1982). The resilience of recursion: a study of a communication system developed without a conventional language model. In Wanner, E. and Gleitman, L., editors, *Language acquisition: the state of the art*, pages 51–77, New York. Cambridge University Press.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Science*, 3(7):272–279.

Kegl, J. and Iwata, G. (1989). Lenguaje de Signos Nicaragüense: A pidgin sheds light on the "creole?". In Carlson, R., DeLancey, S., Gilden, S., Payne, D., and Saxena, A., editors, *Proceedings of the Fourth Annual Meeting of the Pacific Linguistics Conference*, Eugene. Universty of Oregon, Dept. of Linguistics.

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 Briscoe, T., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge University Press.

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

Peters, A. M. (1983). *The units of language acquisition.* Cambridge University Press, Cambridge.

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

Prigogine, I. and Strengers, I. (1984). *Order out of Chaos.* Bantam Books, New York.

Sankoff, G. and Laberge, S. (1973). On the acquisition of native speakers by a language. *Kivung*, 6:32–47.

Sebba, M. (1997). *Contact languages: Pidgins and Creoles.* Palgrave, New York.

Senghas, A. and Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, 12(4):323–328.

Senghas, A., Kita, S., and Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305(5691):1779–1782.

Smith, A. D. M. (2005). The inferential transmission of language. *Adaptive Behavior.* In press.

Smith, K., Brighton, H., and Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in Complex Systems*, 6(4):537–558.

Smith, K. and Hurford, J. R. (2003). Language evolution in populations: extending the iterated learning model. In Banzhaf, W., Christaller, T., Ziegler, J., Dittrich, P., and Kim, J. T., editors, *Advances in Artificial Life: Proceedings of the 7^{th} European Conference on Artificial Life*, Lecture Notes in Computer Science: Lecture Notes in Artifical Intelligence, pages 507–516. Springer-Verlag, Heidelberg.

Steels, L. (1996a). Emergent adaptive lexicons. In Maes, P., editor, *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulating Adaptive Behavior*, Cambridge Ma. The MIT Press.

Steels, L. (1996b). Perceptually grounded meaning creation. In Tokoro, M., editor, *Proceedings of the International Conference on Multi-Agent Systems*, Menlo Park Ca. AAAI Press.

Steels, L. (2004). Constructivist development of grounded construction grammars. In Daelemans, W., editor, *Proceedings Annual Meeting of Association for Computational Linguistics.*

Steels, L. and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language. a case study for colour. *Behavioral and Brain Sciences.* In press.

Steels, L., Kaplan, F., McIntyre, A., and Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In Wray, A., editor, *The Transition to Language*, Oxford, UK. Oxford University Press.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74:209–253.

van Zaanen, M. (2000). ABL: Alignment-based learning. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING).*

Vogt, P. (2003). THSim v3.2: The Talking Heads simulation tool. In Banzhaf, W., Christaller, T., Dittrich, P., Kim, J. T., and Ziegler, J., editors, *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL).* Springer Verlag Berlin, Heidelberg.

Vogt, P. (2005a). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence.* In press.

Vogt, P. (2005b). Meaning development versus predefined meanings in language evolution models. In *Proceedings of IJCAI-05.* In press.

Vogt, P. (2005c). Stability conditions in the evolution of compositional languages: issues in scaling population sizes. In *Proceedings of the European Conference on Complex Systems, ECCS'05.* In press.

Vogt, P. and Divina, F. (2005). Language evolution in large populations of autonomous agents: issues in scaling. In *Proceedings of AISB 2005: Socially inspired computing joint symposium.*

Vogt, P. and Smith, A. D. M. (2005). Learning colour words is slow: a cross-situational learning account. *Behavioral and Brain Sciences.* In press.

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

Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology.* Addison-Wesley, Cambridge, MA.