*Adaptive Behavior*

# Verifying Theories of Language Acquisition Using Computer Models of Language Evolution

Paul Vogt[1], Elena Lieven[2]
[1] *Tilburg centre for Creative Computing, Tilburg University, The Netherlands*
[2] *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

This article presents a dense database study of child language acquisition from a usage-based perspective and a new analysis of data from an earlier study on simulating language evolution. The new analysis is carried out to show how computer modeling studies can be designed to generate predictions (results) that can be compared quantitatively with empirical data obtained from the dense database studies. Although the comparison shows that the computer model in question is still far from realistic, the study illustrates how to carry out agent-based simulations of language evolution that allow quantitative verification of predictions with empirical data to validate theories on child language acquisition.

**Keywords** usage-based approach · dense database studies · early child grammar · traceback methods · language games

## 1 Introduction

Studies on human language evolution and language acquisition are intimately connected. It is generally accepted that one of the major evolutionary adaptations—if not *the* major adaptation—of *Homo sapiens* (or perhaps their immediate ancestors) has to do with our capacity to acquire (and use) language. It is unclear, however, what that adaptation was or whether there was more than one. Studies of child language acquisition can inform us about the socio-cognitive mechanisms of how modern humans acquire (usually) an existing language, thus providing the target toward which we have evolved.

Theories that try to explain these socio-cognitive mechanisms are, however, often hard to verify because of at least two reasons: a) it is difficult to observe and experiment with the whole complex process of language acquisition empirically, and b) the theories tend to be descriptive rather than formal, which makes it difficult to generate hard quantitative predictions that can be verified empirically (Vogt & de Boer, 2010). We believe that computational modeling provides a useful tool regarding this second point. Once a theory about first language acquisition has been verified, our knowledge concerning language evolution would also immediately advance, not the least because new reliable predictions could then be made about language evolution.

One influential theory regarding language evolution is that humans have evolved a language acquisition device (LAD) that encodes a universal grammar (UG), either as a spandrel (Chomsky, 1980) or through natural selection (Pinker & Bloom, 1990). According to this theory, our brains have evolved language-specific areas ready for acquiring the complex grammatical structures of language, as specified in UG. One of the pri-

*Correspondence to*: Paul Vogt, Tilburg centre for Creative Computing, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. *E-mail*: science@paul-vogt.nl. *Tel.*: +31 13 4668118, *Fax*: +31 13 4662892


(2010), Vol 18(1): 21–35.
DOI: 10.1177/1059712309350970

mary reasons for positing a UG is the poverty of the stimulus argument, which assumes that the input to children is too limited to allow them to acquire the input language with adult competence (Chomsky, 1956). However, recent research on child language has suggested that the "poverty" of the stimulus has to be revised in two ways. Firstly, if the child's processing of the input is seen as probabilistic rather than as the extraction of all-or-nothing rules, the logical problem of generating over-general grammars becomes much more tractable (see MacWhinney, 2004, and commentaries). Secondly, there is increasing evidence that children's learning of syntactic constructions is extremely closely related to characteristics of the input (Freundenthal, Pine, Aguado-Orea, & Gobet, 2007; Lieven, Behrens, Speares, & Tomasello, 2003; Rowland, 2007; Theakston, Lieven, Pine, & Rowland, 2005).

An alternative view is that the grammars of human languages have evolved with non-language specific brains, but as the result of cultural evolution (Bates, 1979; Christiansen & Chater, 2008; Tomasello, 1999). According to this usage-based approach (Croft, 2000; Langacker, 2000; Tomasello, 2003), languages are learnt through general pattern recognition mechanisms, allowing humans to construct their internal languages on the basis of actual speech events. When languages are then repeatedly transmitted from one generation to the next, their grammatical structures could have evolved culturally from initially holistic or idiomatic languages to accommodate the cognitive learning mechanisms of humans, thereby improving learnability (Deacon, 1997; Kirby, Dowman, & Griffiths, 2007; Vogt, 2007; Wray, 1998). Although this view does not presume a language-specific adaptation to acquire the universal tendencies of grammar, none of its proponents deny there must have been one or more biological adaptations within the human lineage that gave rise to us becoming language users. Such adaptations, however, are rather sought in, for instance, the development of species-specific symbolic skills (Deacon, 1997) and/or adaptations involving the understanding that other individuals have similar intentions (Tomasello, 1999).

In both views, the nature of language acquisition mechanisms is a central issue. Studies on child language acquisition are, therefore, crucial to increase our knowledge regarding its nature. Rather than studying the dichotomy between nature and nurture, we will adopt the usage-based approach and seek support from both child language studies and computational modeling. In particular, we attempt to analyze data obtained from Vogt's (2005a, 2005b) computational model that implements the usage-based approach and compare these with empirical data obtained from dense corpora of child language to explore children's productions of multiword utterances (such as Dąbrowska & Lieven, 2005; Lieven, 2006; Lieven et al., 2003; Lieven, Salomo, & Tomasello, 2009).

In language evolution research, there have been a number of studies in which the results of computer simulations have been compared with empirical data, for example, in relation to the emergence of vowel systems (de Boer, 2001) or color names (Steels & Belpaeme, 2005). However, to our knowledge there is no study in which outcomes of computer simulations are quantitatively compared with empirical data on child language development. The study presented in this article is a first attempt, but because of the oversimplifications of the model, the result of the comparison is still rather poor. The purpose of this article, however, is not to verify the usage-based approach or the model, but to illustrate how computer models could be used to generate predictions that can be compared with data obtained from empirical studies—the topic of this special issue (Vogt & de Boer, 2010).

In the next section, research using dense corpora of child language to explore children's production of novel multiword utterances will be presented to illustrate the workings of this approach and provide data that can be compared with results generated computationally. Section 3 will present a novel analysis of the study presented in Vogt (2005b) to illustrate how such a comparison can be achieved. In Section 4 we will discuss the approach, including necessary adaptations and cautions.

## 2 The Usage-Based Approach to Language Acquisition

In contrast to a nativist-linguistic approach to language acquisition which views the child as already in possession of a highly abstract universal grammar through which the particular characteristics of the input language are read and learned, usage-based approaches view the child as learning mappings of sound and meaning as constructions for communication (Dąbrowksa &

Lieven, 2005; Tomasello, 2003). These will initially be fully concrete "chunks" of speech, mapped to child-identified meanings and may be one or more words long. As development proceeds, these constructions become internally analyzed and related to each other in a network ("inventory," Langacker, 2000) of constructions. Constructions become more abstract through the development of schematic slots and more complex through the internal development of these slots and the addition of new slots. Type and token frequencies are seen as being crucial to this process of schematization: high token frequencies will establish a lexically specific chunk, while high type variation within a chunk of speech will lead to the development of a slot. Lexically specific strings may remain accessible in the inventory if they are heard and used frequently enough (Bybee & Scheibman, 1999) but may also become internally analyzed. Note that children's linguistic creativity and capacity for abstraction is not in doubt in this approach: what is at issue is the scope of this creativity, the precise basis of the abstractions and how these factors change with development.

There is evidence for considerable repetitiveness in the speech that adults address to children. In studies analyzing the first 1–3 words in the utterances of English child directed speech (CDS), it was found that over 86% of the utterances could be accounted for by a mean of 143 frames per mother (strings that had occurred more than four times in one's mother's speech; Cameron-Faulkner, Lieven, & Tomasello, 2003; Stoll, Abbot-Smith, & Lieven, 2009). Stoll et al. (2009) also found similar results for German and Russian CDS, though typological differences between the languages did somewhat lower the repetitiveness of initial strings. There is also evidence that children can retain and reproduce highly frequent strings. The importance of high frequency strings in children's language learning is also supported by work on children's errors (Kirjavainen, Theakston, & Lieven, in press; Rowland, 2007; Theakston & Lieven, 2008; Theakston et al., 2005). For instance, Rowland (2007) showed that children's wh-inversion errors were more likely for wh-aux sequences that did not appear with high frequency in the children's input while those that had a higher frequency were significantly less likely to cause error, even when the frequency of the individual words in the strings was controlled.

Thus high frequency strings are present in the input, are learned, and can protect the child from error. How then might this relate to children's linguistic creativity and linguistic development? Using a method called "traceback," a number of studies have attempted to address this question by identifying novel utterances and attempting to search for their antecedents in prior parts of the corpus (more detailed accounts can be found in Dąbrowska & Lieven, 2005; Lieven et al., 2003; Lieven et al., 2009).

## 2.1 Constructing Novel Utterances: The "Traceback" Method

In the traceback method we have used corpora recorded from four 2-year-old children with their caretakers. The corpora were collected over a 6-week period for 5 hours a week, yielding 28–30 hours of recording. We divide each corpus into a "test corpus" consisting of the last 2 hours of recording and a "main corpus" consisting of the previous 26–28 hours. We then identify all multiword utterance types in the test corpus and attempted to identify the "component units" from the main corpus that could have been used to construct them. In attempting to establish how close a child's utterance is to what has been said before, there is the issue of whether to look only at the child's previous corpus or whether to include the input. In some previous traceback studies the input was included on the grounds that it is from the input that the child is learning (Dąbrowska & Lieven, 2005; Lieven, 2006; Lieven et al., 2003). The argument for tracing back only to the child's own corpus is also strong: we then know that the string is (or rather was, when it was uttered) part of the child's linguistic representation and in Lieven et al. (2009), therefore, the children's utterances were traced back only to their own previous utterances.[1] In the present study, however, we decided to include the caretaker input in the main corpus since this is closer in spirit to the models discussed below.

## 2.2 Method

We identified all multiword utterance types in the test corpus and then attempted to find matching component units in the main corpus. This was done using a program called Autotracer[2] which produced a list of strings in the main corpus that contained any overlapping lexical material with the target utterance in the test corpus. To be considered as a component unit, each string had to occur at least twice in the main cor-

**Table 1** Types of slots.

| Type of slot | Example utterances | | Schema with slot |
|---|---|---|---|
| REFERENT | CHI | More choc+choc **on there**. | **REFERENT on there** |
| | CHI | Bow-'s food **on there**. | |
| PROCESS | CHI | **I want to** get it. | **I want to PROCESS** |
| | MOT | And **I want to** talk to you about the park. | |
| ATTRIBUTE | CHI | Pilchard there **he's** hungry @sc toast. | **He's ATTRIBUTE** |
| | CHI | **He's** upside+down. | |
| LOCATION | CHI | **I sit** on my Mummy-'s bike. | **I sit LOCATION** |
| | CHI | **I sit** there. | |
| DIRECTION | CHI | **Going** under bridge. | **Going DIRECTION** |
| | CHI | **Going** down. | |
| POSSESSOR | INV | This is my **favorite**. | **POSSESSOR favorite** |
| | MOT | Yeah it's your **favorite** that one, isn't it? | |
| UTTERANCE | CHI | Open the door **Mummy**. | **UTTERANCE Mummy** |
| | INV | There's the cook **Mummy**. | |

pus. Two types of component units were defined: "fixed strings" and "schemas with slots." Fixed strings were fully lexically specific while schemas contained slots. The strings that filled slots could be either fixed strings (single or multiword) or other schemas with slots, however, slots could only be identified in schemas if they were semantically coherent, and the single or multiword strings that filled the slots also had to match the semantics of the slot. Research assistants used the program output to identify semantically coherent strings and slots and to derive the potential tracebacks. This criterion of "semantic coherence" resulted in the following slots being identified: REFERENT, PROCESS, ATTRIBUTE, LOCATION, DIRECTION, POSSESSOR, and UTTERANCE (see Table 1).

Once all potential component units in the main corpus had been identified, a derivation of the novel utterance was attempted using three operations: SUPERIMPOSE, SUBSTITUTE, and ADD. SUPERIMPOSE and SUBSTITUTE place a component unit into the matching slot of a schema. In a SUPERIMPOSE operation, the component unit placed in the slot overlaps with some lexical material of the schema, while in a SUBSTITUTE operation it just fills the slot (see Table 2). ADD places component units at one or other end of an utterance. It had to be syntactically and semantically possible for these units to go in either order so, for instance, conjunctions such as *and* and *but* could not be used in an ADD operation. Vocatives (e.g., *Mummy*, *Daddy*, and Proper names) and adverbials such as *now* and *then* were the primary examples (see Table 3). For some utterances, a number of alternative derivations from component units were possible. This was particularly the case for vocatives that could be analyzed either as consisting of an UTTERANCE slot plus a vocative or as an attested component unit with the vocative attached using the ADD operation. Since we were interested in seeing how far an approach using schemas and slots within them could account for the target utterances in the test corpora, we always used substitution into the utterance slot rather than an add operation if both were possible. In addition, the following rules were used to reduce the tracebacks to the minimum number of operations:

- the longest possible schemas were used;
- the slots were filled by the longest available units;
- the minimum number of operations was taken.

Table 2 gives an example of a single operation traceback using a SUPERIMPOSE operation. Although there is an exact match for the target utterance, *I can't open*

**Table 2** Target utterance: *I can't open it* (Annie 2;0).

| Component units | | Frequency in main corpus | | |
|---|---|---|---|---|
| | | CHILD | | CARETAKER(S) |
| I can't open it | | 1 | | 0 |
| I can't PROCESS it | | 17 | | 6 |
| can't open it | | 2 | | 1 |
| Step | Component unit | Operation | Filler of slot | Result |
| 1 | I can't PROCESS it | SUPERIMPOSE (PROCESS) | can't open it | I can't open it |

**Table 3** *And that's Emma outside* (Annie 2;0).

| Component units | | Frequency in main corpus | | |
|---|---|---|---|---|
| | | CHILD | | CARETAKERS |
| UTTERANCE outside | | 0 | | 8 |
| and that's REFERENT | | 0 | | 14 |
| Clare | | 4 | | 5 |
| outside | | 12 | | 19 |
| Step | Component unit | Operation | Filler of slot | Result |
| 1 | UTTERANCE outside | SUBSTITUTE (UTTERANCE) | And that's REFERENT | And that's REFERENT outside |
| 2 | and that's REFERENT outside | SUBSTITUTE (REFERENT) | Emma | And that's Emma outside |

*it,* in the main corpus, the criterion of two utterances for an exact match means that it cannot be used. Instead we find 17 examples in the child's utterances of *I can't PROCESS it* and six from the adults plus three of the fixed string *can't open it*. This allows *can't open it* to be superimposed onto *I can't PROCESS it* (it is a SUPERIMPOSE rather than a SUBSTITUTE because the matching string shares material with the schema, the lexemes to either side of the slot, *can't* and *it*). This gives the target utterance in a one-operation traceback.

In Table 3 we give an example of the use of an UTTERANCE slot. This involves two operations of SUBSTITUTE, one of the schema *and that's REFERENT* into the schema *UTTERANCE outside*, the other of the fixed string Clare into the *REFERENT* slot of *and that's REFERENT*.

The traceback analysis was carried out, after extensive training, by two research assistants. For reliability, 20% of all tracebacks were coded twice. Agreement was high (kappa = 0.89)

## 2.3 Results

Figure 1 gives the overall data for the four children's corpora at 2 years. The children are ordered from left to right in increasing mean length of utterance (MLU) calculated in words over the whole corpus (Brian,

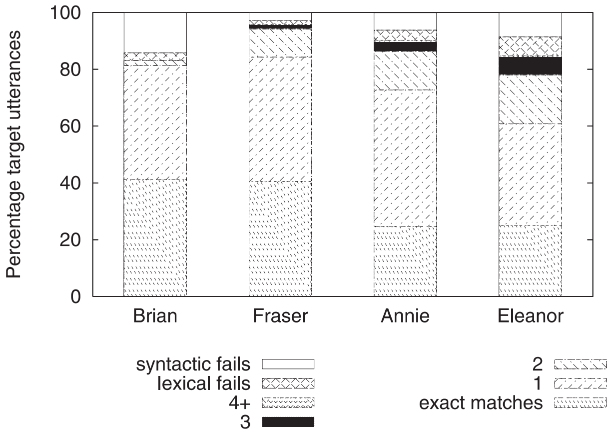


**Figure 1** Number of operations required to trace back target utterances to the main corpus at 2;0.

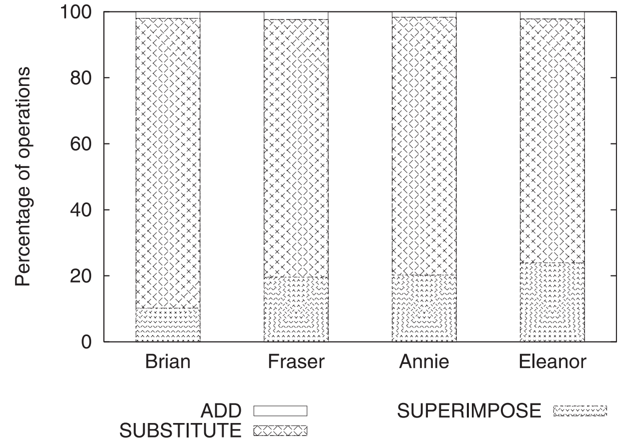


**Figure 2** Types of operations.

1.65, Fraser 1.8, Annie, 2.19, Eleanor, 2.22). Between approximately 25% (Annie and Eleanor) and 40% (Brian and Fraser) of novel utterances in the test corpora have actually occurred at least twice in exactly the same form in the main corpus (exact matches) and a further 36% (Eleanor) to 48% (Annie) require only a single operation to arrive at a successful traceback. The proportion of multi-operation tracebacks (requiring 2, 3, and 4+ operations) increases with increasing MLU from 1.8% for Brian to 24% for Annie and there are between 4% (Annie) and 17% (Brian) of utterances that cannot be traced back (fails). SUBSTITUTIONs formed the vast majority of operations (from 74–88%) and the number of ADD was only 2% for all tracebacks (see Figure 2).

Thus at 2;0 most of these children's utterances are either exact repeats of something they have said or heard before or require one single operation to arrive at a successful traceback. Note that with increasing MLU, the proportion of exact matches goes down and the number of multi-operation tracebacks goes up. In Dąbrowska and Lieven (2005), we found similar results when we compared tracebacks of syntactic questions at 2;0 and 3;0. This is, of course, directly related to the increase in the length of utterances in development, but it also reflects a greater productivity and less reliance on exact repetition with development.

It is interesting to note that by far the largest proportion of slots in all tracebacks is for REFERENTS (60–89%). The proportion of non-referent slots increases with MLU and, in particular, while Brian has almost no PROCESS slots and no ATTRIBUTE slots, the proportion of these is higher for Fraser and even higher for Annie and Eleanor (see Figure 3). This fits with research suggesting that, at least for English-speaking children, the category of nouns is abstracted earlier than that of verbs or other categories. For instance, children can substitute novel object names into frames from about 1;9 (Tomasello, Akhtar, Dodson, & Rekau, 1997). In Lieven et al. (2009) we analyze the strings that fill the REFERENT slots and show that, as MLU develops, the proportion of bare nouns reduces and the proportion of NPs with determiners increases, as does the range of determiners. We suggest that once children have abstracted a function to form mapping of REFERENT to Noun, they can then

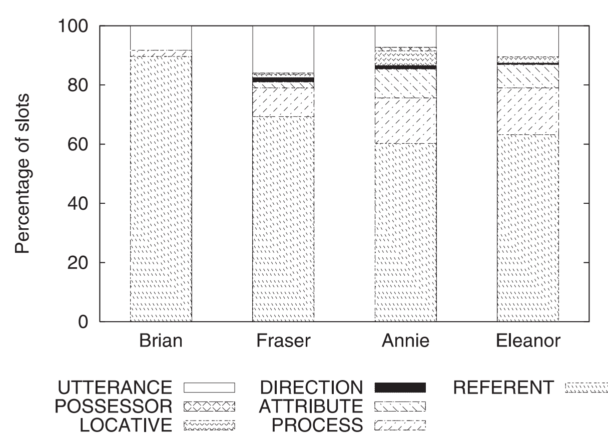


**Figure 3** Types of slots.

start to build the linguistic constituent of Noun Phrase. Thus, in our approach, grammatical structure is emergent rather than pre-given.

In summary, we were able to trace back a very large proportion of these 2-year-olds' utterances in the test corpus either to exact repetitions in the main corpus or to schemas that required only one substitution into a slot. The vast majority of these substitutions were of single nouns. With developing MLU, the range of other slots developed, as did the variety of REFERENT strings.

### 2.4 Discussion

Despite the fact that we estimate that we are only capturing 7–10% of the children's waking lives, we are able to account for 61–94% of their novel utterances in the test corpus in terms of exact repetitions or a single operation. Our main interest here was to see whether children's utterances could be related to lexically specific units that we know they have previously encountered. To know whether they actually *used* this particular method of constructing their utterances, we would need to know much more about how such units are retrieved and processed. Some support for the idea that children are learning chunks rather than just assembling utterances from individual words is provided by a recent study of Bannard and Matthews (2008). This shows that children were more fluent and made fewer errors in four-word strings with high frequencies in CDS than with strings in which the individual words were matched for frequency but the strings as a whole were of lower frequency. Once a chunk has been learned, the idea is that slots will be formed if there is partial repetition of lexically specific material but also type variation in some part of the string. Although there is relevant work in historical linguistics as to the relation between type and token frequencies that leads either to entrenchment of the whole string or to the formation of slots (Bybee, 1995), this is an important topic for future research in child language studies (Bannard & Lieven, 2009).

The fact that the proportion of exact matches goes down with development is, we would suggest, related to the increase in the child's repertoire of schemas with slots which gives rise both to greater productivity and expressiveness. The very low proportion of ADD operations is partly to the result of our use of the UTTERANCE slot, but this only affected a relatively small number of tracebacks: the largest proportion of UTTERANCE slots (16%) was found in Fraser's tracebacks and most included *Mummy* as the vocative. More interestingly, to ADD two strings required, under our definitions, that they could go in either order, whereas in using SUBSTITUTE or SUPERIMPOSE, the order was already defined, as was the semantic content of the slot. We tentatively suggest that this could provide English-speaking children, who clearly hear a lot of linear repetition in the speech of their environment, with an efficient means of expressing the form-meaning mappings that they wish to communicate. Whether an operation like ADD would be more significant for tracebacks involving languages with freer word order, is an interesting question.

## 3 Simulating the Emergence of Compositionality With a Usage-Based Approach

If simulations can be carried out with a model that implements the socio-cognitive mechanisms proposed in the usage-based approach, then the results are predictions of the theory that should closely match the findings found in the dense database studies. In this section we will analyze a corpus from Vogt's (2005b) study and compare the results with the findings from child language acquisition as presented in the previous section. No attempt has been made to improve the model such that the results closely match the empirical findings. As mentioned, the purpose is not to verify Vogt's model with respect to the data, but merely to illustrate how a proper comparison can proceed.

Vogt's (2005b) study was concerned with the emergence of *compositionality*. Compositionality means that parts of an expression have a functional relation to the parts of its meaning. For example "John loves Mary" is compositional in that each element has a distinct meaning and the meaning of the whole sentence is a particular combination of these meaning parts. From a usage-based perspective, we assume that compositional structures require a schema-like representation with slots. In contrast to compositionality, there is no part in the expression "bought the farm" that refers to any part of its meaning *died*. Such expressions are "fully lexically specific" or *holistic*.

## 3.1 The Model

The model is an adaptation of Kirby's (2001) iterated learning model with which it was shown that initially holistic languages can evolve culturally into compositional ones when the language is iteratively transmitted over generations through a transmission bottleneck (such that children only observe a subset of the target language). Vogt's model differs in a number of ways (see, Vogt, 2005a, 2005b, 2007, for discussions). For the current study, it is most important to mention that Vogt's model allows communication in all directions, whereas Kirby's model only allows vertical communication (i.e., from adults toward children), as a result of which the transmission bottleneck is implicit during the children's development.

The model implements a simulation of the Talking Heads experiment (Steels, Kaplan, McIntyre, & Van Looveren, 2002) in which a population of agents plays a large number of *guessing games* to develop a language that allows the population to communicate about their world, which contains a number of colored geometrical shapes. It is impossible to present all details of the model in this article; the interested reader is referred to Vogt (2005a, 2005b). For each guessing game, two agents are selected from the population arbitrarily: one speaker and one hearer. Both agents are presented eight objects (the *context*) randomly sampled from the virtual world that contains a total of 120 objects (12 colors × 10 shapes). Each agent individually tries to categorize the perceptual features of each object in the context using a method based on the discrimination game (Steels, 1996), whose details are irrelevant to the scope of this article. Suffice to say that each object is categorized such that it is distinctive from all other objects in the context. If distinctive categorization fails, a new category is constructed for which the object's perceptual features serve as exemplars. (Note that initially each agent has no categories at all; these are all constructed when needed by these discrimination games.) Categories are represented as prototypical points in a 4-dimensional space, each dimension relating to a perceptual feature, which are the red, green, and blue components of the RGB color space and a shape feature.

The 4-dimensional categories will be taken as the meaning of a whole expression. It has been shown that during the process of cultural evolution, the 4-dimensional meanings can be restructured to form meanings of lower dimensions. This way, *conceptual spaces* (Gärdenfors, 2000) are constructed, which could be interpreted as linguistic categories. The language learning mechanisms that we explain shortly guides this restructuring.

| | | |
|---|---|---|
| 1 | S → greensquare/(0,1,0,1) | 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 4** This example grammar contains rules that rewrite a non-terminal (A, B, or S) into a terminal form-meaning pair (1, 3, and 4) or into a compositional rule, that is, schema, that combines different non-terminals, that is, slots (2). The meanings are 4-dimensional vectors, where the first three dimensions relate to the RGB color space (rgb) and the fourth one relates to the shape feature (s). The question marks are wild-cards, which may be substituted by any other category in that dimension. Each rule has a score that indicates its effectiveness in past guessing games. Only expressions of two constituents are allowed in this grammar.

Once the agents have categorized the objects in the context, the speaker selects one object at random as the topic of the communication. This agent then searches its grammar for ways to produce an expression that conveys the topic's meaning. The grammar (Figure 4) consists of simple rewrite rules that associate forms with meanings either holistically (e.g., rule 1) or compositionally (e.g., rule 2 combined with rules 3 and 4). The grammar may be redundant in that there may be rules that compete to encode or decode an expression (cf., Batali, 2002; De Beule & Bergen, 2006). The speaker searches for those (compositions of) rules that match the topic's meaning and if more than one are found, he selects the rule that has the highest rule score. Sometimes when the speaker produces a compositional construction, this may be a novel one. The operation to construct such a novel utterance is similar to the SUBSTITUTE operation referred to in the previous section, but may also be an EXACT MATCH of a previously heard expression. If the speaker fails to produce an expression this way, a new form is invented as an arbitrary string and is associated with the topic's whole meaning (INCORPORATE) or—if a part of the meaning matches some non-terminal rule—with the rest of this meaning (EXPLOIT).

In turn, the hearer tries to interpret the expression by searching its own grammar for (compositions of) rules that match both the expression and a category relating to an object in the current context. If there is more than one such rule, the hearer selects the one with the highest score, thus guessing the object intended by the speaker. The hearer then points to this object, and if this is the object intended by the speaker, the speaker acknowledges success; otherwise, the speaker points to the topic allowing the hearer to acquire the correct meaning. If the guessing game was successful, both the speaker and hearer increase the scores of the rules they used and lower the scores of those rules that compete with the used rules. (When the hearer thus interpreted a—to him—novel utterance, we say he used the SUBSTITUTE operation.) If the game has failed, the scores of used rules are lowered and the hearer acquires the proper association between the heard expression and the topic's meaning. To this end, the hearer tries the following three induction steps until one step has succeeded:

1. EXPLOIT: If a part of the expression can be decoded with a part of the topic's meaning, the rest of the expression is associated with the rest of the meaning. For instance, if the hearer of the grammar shown in Figure 6 hears the expression "redcircle" meaning (1,0,0,.5), the part "red"-(1,0,0,?) can be decoded, so the hearer adds rule B→circle/(?,?,?,.5) to its grammar.
2. CHUNK: If the above failed, the hearer searches its memory, where it stores all heard or produced expression–meaning pairs, to see if there are instances that are partly similar to the expression–meaning pair just heard. If some similarity can be found, the hearer will break-up the expression–meaning pairs containing the similarities—following certain heuristics, thus forming new compositional rules. Suppose, for instance, the hearer had previously heard the expression–meaning pair "greensquare"-(0,1,0,1) and now hears the expression–meaning pair "yellowsquare"-(1,1,0,1). The hearer can then break up these pairs based on the similarity "square"-(?,1,0,1), thus forming rules S→C/r D/gbs, C→green/(0,?,?,?), C→yellow/(1,?,?,?) and D→square-(?,1,0,1).[3]
3. INCORPORATE: If the above adaptations both fail, the heard expression–meaning pair is incorporated holistically, leading to a new rule such as S→yellowcircle/(1,1,0,.5).

At the end of these steps, the hearer performs a few post-processes to remove any multiple occurrences of rules and to update the grammar such that other parts of the internal language relates more consistently to the new knowledge. Full details of the model are found in (Vogt, 2005a, 2005b).

## 3.2 Results

In Vogt (2005b) it was shown that initially holistic languages can evolve into compositional ones, because of children's creativity in producing novel constructions by—in this model—applying the SUBSTITUTE mechanism described in the previous section. The reason for this creativity is that when children need to communicate about some novel referent (i.e., an object they have not seen before), they may be able to use a schema (rule in this model) and SUBSTITUTE one or more slots using lexical components (parts of phrases) they have heard previously. Otherwise, they may have to invent a new word, which can then either be used to refer to a part of the sentence's meaning using the EXPLOIT operation or to the whole meaning using INCORPORATE. It is not hard to see that if INCORPORATE would occur too frequently, a language can change very rapidly, which is the main reason compositionality does not emerge as a stable system when transmitted vertically without a transmission bottleneck (Vogt, 2005a).

We have replicated one of the conditions from the study presented in Vogt (2005b) with exactly the same initial conditions. The simulations were set up with a population of six agents (three adults and three children) and were run for 10 iterations of 6,000 guessing games each. In each game, the participating agents were selected at random with an equal chance that a speaker/hearer was an adult or a child (cf. one condition in Vogt, 2005b). At the end of each iteration, the adults were removed, the children became adults and new children were included. The simulations were repeated 15 times to generate 15 corpora of linguistic interactions. In the remainder of this section, we will analyze the constructions of the children from the final iteration. Prior to this iteration, the adults had acquired their language with a communicative success of 79.3% on average (with standard deviation of 5.5%), while 80.3 ± 6.3% of their constructions were compositional. The whole population at the end of the simulation achieved on average 79.6 ± 5.5% communicative success and 78.8 ± 5.9% compositionality.

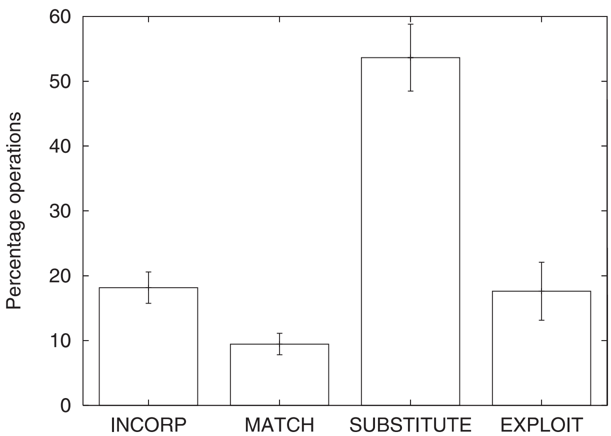


**Figure 5** The percentage of the operations EXPLOIT, SUBSTITUTE, INCORPORATE, and EXACT MATCH used to construct novel utterances as speakers. The results indicate averages over all children in the final iteration of all 15 runs with standard deviations.

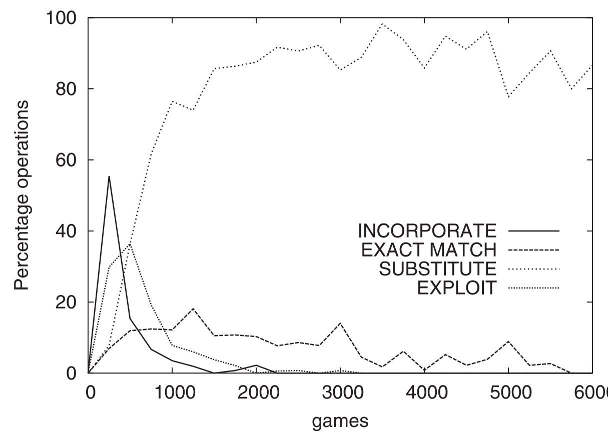


**Figure 6** The percentages of operations used to produce novel utterances over time per 250 games. The results show the averages for all children in the final iteration of the 15 runs.

By applying the traceback method on the generated corpora, we measured the relative frequencies with which the operations EXACT MATCH, SUBSTITUTE, EXPLOIT, and INCORPORATE were used by the three children of the final iteration. Figure 5 shows the relative frequencies (averaged over all children in all runs) with which these operations have been used during the agents' childhoods for *producing* novel utterances as speaker. The results show that for producing (i.e., encoding) novel utterances, SUBSTITUTE was the most frequently used operation for all agents (on average 53.6 ± 5.2%), followed by INCORPORATE and EXPLOIT (18.2 ± 2.4% and 17.6 ± 4.4% respectively), while EXACT MATCH (9.5 ± 1.7) occurred least frequently. It is of course hard to compare these results with those presented in the previous section, but the simulation confirmed the high occurrence frequency of the SUBSTITUTE operation.

Naturally, it is relatively straightforward to plot the relative frequencies with which the various operations are used over time during development. Figure 6 shows the average relative frequencies of operations used to produce novel utterances during the childhood period of the agents across the simulations over time (the frequencies were measured per time window of 250 guessing games). It is interesting to see that during development there is an initial period where most constructions are based on INCORPORATE (creating new utterances) and EXPLOIT, but that SUBSTITUTE soon contributes to, on average, around 90% of the novel constructions after, on average, 2,000 games. Looking at individual runs, it was found that actually from game 3,500 onwards most simulations yielded relative frequencies of 100% SUBSTITUTE operations. The reason that the average is lower than 100% is that the number of novel constructions diminish later on in the child's development to the extent that in some of the simulations no novel constructions were found over the period measured. The results predict that in a longitudinal corpus of child language, one would find early in the child's development fewer SUBSTITUTE operations than, for instance, EXACT MATCH operations, whereas the latter tends to diminish later on in favor for the SUBSTITUTE operations. This is consistent with findings from the dense database studies, although the amount of novel constructions tends to increase instead of decrease. The reason that this increase does not occur in the simulation is that the language size is highly limited, so no novel constructions are found after the children in the simulation covered all possible meanings.

## 3.3 Discussion

It is clear that the results of these simulations are hard to compare with the results presented in the previous section for the construction produced by real children, because the model here is too limited (e.g., only one- or two-word construction are allowed and the inputs to children are also only one- or two-word constructions). Nevertheless, this study illustrates how the results of a computer simulation could be compared with real child language data and vice versa.

## 4 General Discussion

In this article we have presented an empirical study of child language acquisition based on applying a traceback procedure to dense child language corpora. We then presented a study that applies this traceback procedure to an artificial language corpus generated by an agent-based model that simulates aspects of language evolution. It is still very much work in progress—in fact this is our first attempt to compare quantitatively empirically obtained corpora of child language with artificially generated corpora—but this study is meant to illustrate how one can incorporate empirical data to generate and verify predictions using computational modeling (cf. Vogt & de Boer, 2010).

Of course, corpus-based studies typical of computational linguistics are also very useful for investigating certain cognitive mechanisms underlying language acquisition, because they can deal with real empirical input to children (e.g., Bannard, Lieven, & Tomasello, 2009; Borensztajn, Zuidema, & Bod, 2009; Freudenthal et al., 2007; Monaghan, Christiansen, & Chater, 2007; Redington, Chater, & Finch, 1998) or with natural written or spoken languages (e.g., Bod, in press; Daelemans & van den Bosch, 2005). However, such methods lack the bidirectional social interactions crucial to language use, thus they fail to capture the incremental effect that children's productions in dialogues can have on language acquisition. Agent-based models do simulate such interactions.

Many agent-based models studying the emergence of compositionality and other syntactic phenomena have adopted the usage-based or constructive approach (Batali, 2002; Brighton, 2002; Kirby, 2001; Steels & De Beule, 2006; Vogt, 2005a, 2005b). Yet, these models differ substantially in how they are implemented. For instance, Batali's (2002) model incorporates schemas of binary trees that fit into each other, while Brighton's (2002) model constructs finite state automata based on finding minimum description lengths. Kirby's (2001) model uses rewrite rules as its basis, and Steels & De Beule's (2006) model produces (fluid) construction grammars anchored in the agents' interaction with the real world. The models of Batali, Brighton, and Kirby assume predefined semantics. In Vogt's (2005a, 2005b) model, which is an adaptation of Kirby's model, the semantics are anchored in the agents' interaction with a virtual world and where the grammars become networks of rules that can compete with each other for usage.[4]

These models have in common that they show that relatively complex grammatical structures can emerge through cultural evolution using more (Brighton, 2002; Steels & De Beule, 2006) or less (Batali, 2002; Kirby, 2001; Vogt, 2005a) complicated pattern finding mechanisms, which are assumed to have evolved for general purposes and not specifically for language. These different models are all based on (slightly) differing interpretations of how the usage-based approach can be defined formally in terms of socio-cognitive mechanisms. Yet, all produce a development that—should the simulations be initialized realistically—can be quantitatively compared more or less directly with data obtained from the dense databases of child language corpora in the manner presented here. The more closely the results (predictions that the theory construct) match those found in the empirical studies, the more likely the theory/model is correct.

We reanalyzed Vogt's (2005a, 2005b) model that implements a usage-based approach to language acquisition based on the operations suggested from the empirical studies presented in Section 2. Simulations with this model have shown that when "children" are triggered to produce an utterance regarding a situation they have not covered before, they can use word-meanings describing parts of the situation that they have already acquired. Vogt's (2005b) study has shown that this can trigger the emergence of compositionality in language. In the current study, we have shown that the model also predicts that the main operation involved in this process is SUBSTITUTE. The same operation has also been observed most frequently in the dense database study on child language (Lieven et al., 2009).

A proper verification, however, is still impossible, because the current model does not yet implement other operations such as ADD and SUPERIMPOSE, nor is the agents' virtual world sufficiently complex to accommodate these operations sensibly. The model therefore does not allow us to make distinctions based on the type of slots, such as REFERENT, PROCESS, ATTRIBUTE, LOCATION, and so forth, as presented in Section 2. It is possible to investigate the evolution of type and token frequencies in this model (see Vogt, 2005a, 2006, for examples), but because of the inability to compare these with the empirical data, we have refrained from doing so. Models that allow for more complex grammatical complexity to emerge (e.g., Batali, 2002; Kirby, 2001; Steels & De Beule, 2006) appear more suitable for such comparisons, though these may still need appropriate adaptations. Simulations with a proper model should statistically yield similar distributions of the use of operations and linguistic units as observed in the dense databases.

When such computer simulations yield statistically similar results to those obtained empirically from child language studies, they positively verify the usage-based theory. (Note that, even if the model is a proper implementation of the theory, a positive verification does not necessarily imply that the theory is correct, but it would provide supportive evidence. Only a negative verification could reject a theory.) Verification could thus shed more light on the nature of language acquisition, but would also provide a target for computational studies on language evolution; that is, it predicts that evolution would have resulted in the socio-cognitive mechanisms proposed in this theory. Moreover, when such verified computer models are used to investigate other aspects of language evolution (e.g., the dynamics of language change), these models are likely to produce more accurate predictions than studies carried out with models that have no empirical grounding.

Comparison between empirical data and computer simulations, however, should be carried out with caution, because empirical data on language acquisition is noisy, while computer models are still crude abstractions of reality. Empirical data is noisy because transcriptions can have errors, experiments may have a wrong set up to begin with, and—most of all—observational and experimental data are capturing only a small portion of a child's language development. Although dense data studies, such as those presented in Section 2 are bridging this gap, at 5 hours of recording a week, they still only capture an estimated 7–10% of a child's waking lives during a limited period. Moreover, such dense data studies, as well as many other longitudinal studies are conducted with very few subjects so that statistical validation is unreliable. The Human Speechome project (Roy et al., 2006) is a major step forward in density, but this is an $n = 1$ study.

Contemporary agent-based models that simulate language evolution tend to model highly simplified languages (e.g., only lexicons or highly simplified grammars), so that the input to simulated children tends to have much lower complexity than the language spoken to human children. Multi-agent simulations in complex environments using empirical data (e.g., from corpora such as CHILDES, MacWhinney, 2000, or those discussed in this article) to simulate the input to children would make the models considerably more realistic. More generally, we believe that in order to improve the quality of predictions generated through computational studies on language evolution, more empirical data obtained from child language research should be incorporated as initial parameter settings into the computer models, which thus improves the ability to verify the model with empirical data reliably (see also, Vogt & de Boer, 2010). Examples of data that can be incorporated include the frequency with which joint attention is used, the statistical nature of the input to children, the number of interlocutors and frequency of child directed utterances, and the proportions of child-directed speech versus overheard speech.

To conclude, the current study has illustrated a new way to compare the results of computer models with those obtained from child language studies. It should be clear that agent-based computer models that simulate language evolution are still far too limited to capture the complex processes involved in child language acquisition, not only in their socio-cognitive capacities, but also in the reduced complexity of the input and their limitations of multimodal interactions. Nevertheless, we do believe that further developments in these types of models can have a significant contribution to our understanding of human language acquisition, provided these models will be more reliably developed based on empirical data, such that their results can be better compared with empirical observations of child language. To achieve this we recommend, like Vogt and de Boer (2010), that the interactions between modeler and empiricist be intensified. Only then can these models be used soundly to verify theories of modern human language acquisition, thus advancing our knowledge about language evolution.

## Notes

1 For a comparison of tracebacks which include caretaker utterances to those using only the child's previous utterances see Dąbrowska and Lieven (2005). This is also briefly discussed in Lieven et al. (2009).

2 The program, Autotracer, was designed by Sascha Hoppe under the supervision of Franklin Chang.

3 Note that the CHUNK example does not yield the ideal break up, since it breaks apart the red component of the RGB color space from the blue and green components and the shape feature. Because of the asymmetry of this space with respect to those colors used, the most compact language can only be achieved when breaking apart rgb from s. It has been shown, though, that such mistakes tend to be repaired during the course of development or evolution (e.g., Vogt, 2006).

4 In Kirby's model there is no competition among rules. See Vogt (2007) for a discussion of the consequences of this and other differences between both models.

## Acknowledgments

This article is the result of the Advanced Studies Initiative workshop *Language evolution: Computer models for empirical data* organized by Paul Vogt and Bart de Boer, and sponsored by a generous grant from the Cognition programme of The Netherlands Organisation for Scientific Research (NWO). We thank all participants of this initiative for their highly stimulating discussions, as well as Bart de Boer and two reviewers for providing invaluable comments to improve this article. Paul Vogt is supported by an NWO Vidi research grant. Elena Lieven would like to thank Katya Hummel, Kristin Wolter, and Silke Harbodt for all the hard work they put into the traceback analyses.

## References

Bannard, C., & Lieven, E. (2009). Repetition and reuse in child language learning. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley (Eds.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, functional explanations.* Amsterdam: John Benjamins.

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences*, *106*(41), 17284–17289.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*, 241–248.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge: Cambridge University Press.

Bates, E. (1979). *The emergence of symbols: cognition and communication in infancy*. New York: Academic Press.

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

Bod, R. (in press). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science*.

Borensztajn, G., Zuidema, W., & Bod, R. (2009). Children's grammars grow more abstract with age. *Topics in Cognitive Science*, *1*, 175–188.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–55.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of "don't" in English. *Linguistics*, *37*(4), 575–596.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27,* 843–873.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–558.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, *2*, 113–124.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, *3*, 1–61.

Croft, W. (2000). *Explaining language change: An evolutionary approach.* Harlow: Longman.

Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*(3), 437–474.

Daelemans, W., & van den Bosch, A. (2005). *Memory-based language processing.* Cambridge, UK: Cambridge University Press.

Deacon, T. (1997). *The symbolic species*. New York: W. Norton and Co.

De Beule, J., & Bergen, B. K. (2006). On the emergence of compositionality. In A. Cangelosi, A. D. M. Smith, & K. Smith (Eds.), *Proceedings of the 6th International Conference on the Evolution of Language* (pp. 35–42). Singapore: World Scientific Press.

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

Freudenthal, D., Pine, J., Aguado-Orea, J., & Gobet, F. (2007). Modelling the developmental patterning of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science*, *31*, 311–341.

Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge, MA: MIT Press.

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., Dowman, M., & Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the USA*, *104*(12), 5241–5245.

Kirjavainen, M., Theakston, A., & Lieven, E. (in press). Can input explain children's me-for-I errors? *Journal of Child Language*.

Langacker, R. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language acquisition*. Stanford: SLI Publications.

Lieven, E. (2006). Producing multiword utterances. In B. Kelly & E. Clark (Eds.), *Constructions in acquisition* (pp. 83–110). Stanford, CA: CSLI Publications.

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

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, *20*(3), 481–508.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. ( 2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, *31*(4), 883–914.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*, 259–305.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707–789.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: a powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*(4), 425–469.

Rowland, C. F. (2007). Explaining errors in children's questions*.* *Cognition*, *104*(1), 106–134.

Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B. et al. (2006). The Human Speechome Project. In *Twenty-eighth Annual Meeting of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum Associates.

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

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language. a case study for colour. *Behavioral and Brain Sciences, 28*, 469–529.

Steels, L., & De Beule, J. (2006). Unify and merge in fluid construction grammar. In P. Vogt et al., (Eds.), *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication* (pp. 197–223). Berlin: Springer.

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: Oxford University Press.

Stoll, S., Abbot-Smith, K., & Lieven, E. (2009). Lexically restricted utterances in Russian, German and English child directed speech. *Cognitive Science, 33,* 75–103.

Theakston, A., & Lieven, E. (2008). The influence of discourse context on children's provision of auxiliary BE. *Journal of Child Language*, *35*, 129–158.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2005). The acquisition of auxiliary Syntax: BE and HAVE. *Cognitive Linguistics*, *16*, 247–277.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

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

Tomasello, M., Akhtar, N., Dodson, K., & Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language*, *24*(2), 373–387.

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. (2006). Cumulative cultural evolution: Can we ever learn more? In S. Nolfi, G. Baldassarre, R. Calabretta, J. Hallam, D. Marocco, J.-A. Meyer, & D. Parisi (Eds.), *From Animals to Animats 9: Proceedings of the Ninth International Conference on Simulation of Adaptive Behaviour*, LNAI (Vol. 4095, pp. 738–749). Berlin/Heidelberg: Springer.

Vogt, P. (2007). Variation, competition and selection in the self-organisation of compositionality. In: B. Wallace, A. Ross, J. Davies, & T. Anderson (Eds.), *The mind, the body and the world.* Exeter: Imprint Academic.

Vogt, P., & de Boer, B. (2010). Language evolution: Computer models for empirical data. *Adaptive Behavior*, *18*, 5–11.

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

## About the Authors

**Paul Vogt** is a senior research fellow at the Tilburg centre for Creative Computing at Tilburg University in The Netherlands. Earlier he held positions at the Vrije Universiteit Amsterdam, the University of Edinburgh, the Universiteit Maastricht, and the Vrije Universiteit Brussel, where he obtained his PhD He holds an MSc in cognitive science and engineering from the University of Groningen. His main interests include modeling language evolution and language acquisition, in particular regarding socio-cognitive and cultural dynamics of symbol grounding. His current research project, CASA MILA, focuses on cross-cultural differences in language acquisition by infants and artificial agents.

**Elena Lieven** is a senior research scientist at the Max Planck Institute for Evolutionary Anthropology, Leipzig and the Director of the Max Planck Child Study Centre at the University of Manchester. She holds a professorship there and an honorary professorship at Leipzig University. She studied experimental psychology at Cambridge and then did her PhD on language development at Cambridge before moving to the University of Manchester. Her principal areas of research involve: usage-based approaches to language development; the emergence and construction of grammar; the relationship between input characteristics and the process of language development; and variation in children's communicative environments. *Address*: MPI-EVA, Deutscher Platz 6, 04103, Leipzig, Germany. *E-mail*: lieven@eva.mpg.de. *Tel.*: 0049-341-3550400, *Fax*: 0049-341-3550-444.