

# OVEREXTENSIONS AND THE EMERGENCE OF COMPOSITIONALITY

PAUL VOGT

*Language Evolution and Computation unit, University of Edinburgh  
40 George Square, Edinburgh EH8 9LL, U.K.  
Computational Linguistics and AI section, Tilburg University  
P.O. Box 90153, 5000 LE Tilburg, The Netherlands  
paulv@ling.ed.ac.uk*

This paper investigates the effect overextensions of words may have on the emergence of compositional structures in language. The study is done using a recently developed computer model that integrates the iterated learning model with the language game model. Experiments show that overextensions due to an incremental acquisition of meanings on the one hand attracts languages into compositional structures, but on the other hand introduces ambiguities that may act as an antagonising pressure.

## 1. Introduction

Over the past decade, a lot of computational models have investigated the emergence of compositional structures in language (for an overview consult, e.g., Briscoe, 2002). Many of these studies have used simulations of multi-agent systems, where the communication system emerges through cultural interactions, individual learning and self-organisation (possibly in combination with the evolution of a LAD). Most of these models have assumed that individual agents (i.e. the individuals of the language community) are ‘born’ with a predefined semantics (e.g., Kirby, Smith, & Brighton, 2004). Naturally, this assumption is not realistic in our human society. This paper focuses on the overextension of meaning, which can occur when the semantics are not predefined, but are developed during an agent’s lifetime.

It is well known that during the process of language acquisition, children go through a phase in which they overextend the meaning of words by using them for inappropriate referents (see, e.g., Clark, 2003). It is unclear what causes this behaviour, but it might be that children cannot yet distinguish among different referents, or that they do not have the proper word for a referent yet. Typically, overextensions occur very early in life and an overextended form can last from one day to several months. Although overextensions are typically considered as a phase relating to the acquisition of word meanings, it is interesting to investigate

if they can have an unexpected (side-)effect.

As a conclusion to a recently studied model on how compositionality can emerge in a simulation in which the semantics develop ontogenetically in individuals, it has been hypothesised that overextensions both can provide a positive attraction towards using compositional structures, and an antagonising pressure against the emergence of compositional structures (Vogt, 2005b, 2005a). This simulation is based on a model that integrates the *iterated learning model* (Kirby et al., 2004) and the *language game* (or guessing game) model of the Talking Heads experiment (Steels, Kaplan, McIntyre, & Van Looveren, 2002).

This paper further explores the hypothesised effect of overextensions on the emergence of compositionality in language. The next section briefly introduces the model. Section 3 then presents some experimental results that test the hypothesis. Finally, Section 4 concludes.

## 2. Grounded iterated learning

Earlier work on the ILM has shown how initially holistic languages can evolve into compositional ones when the language is iteratively transmitted from one generation of individuals to the next, provided the individuals have the appropriate learning mechanisms to discover compositional structures, and the language is transmitted through a *bottleneck*, such that children only learn from a subset of the language (Kirby et al., 2004). One limitation of this earlier work is that all agents start their lives with a predefined semantics.

Similar results were achieved in Vogt (2005b), where the individuals acquire their meanings incrementally as they engage in language games to develop their language. In this model compositionality emerges in the first generation. However, compositionality only remains stable over time when the language is transmitted through a bottleneck, like in the earlier ILMs. When the language is not transmitted through a bottleneck, compositional languages tend to collapse into holistic languages, provided the language is transmitted purely in a vertical direction (i.e. all speakers are adults and all hearers are children).

The model used in these recent studies is implemented in a simulation toolkit of the Talking Heads experiment (Steels et al., 2002), called THSim.<sup>a</sup> In this model, a population of agents tries to evolve a simple language by which they can describe geometrical coloured objects presented to them. The agents achieve this by engaging in a series of guessing games, which are played by two agents – a speaker and a hearer – selected from the population. In the model discussed here, all speakers are selected from the adult population and all hearers from the child population; thus the language is transmitted vertically like in most ILMs. The remainder of this section summarises the model very briefly; for a detailed explanation, the reader is referred to Vogt (2005b).

---

<sup>a</sup>THSim is available at <http://www.ling.ed.ac.uk/~paulv/thsim.html>.

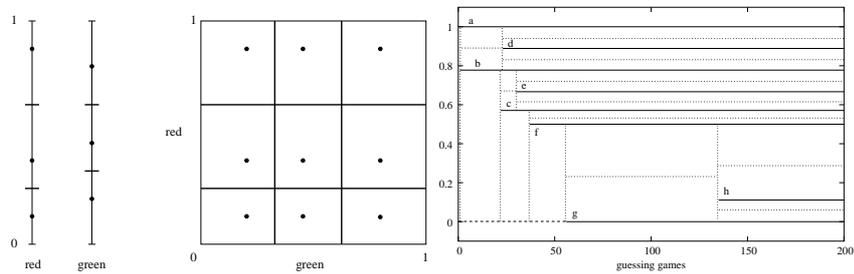


Figure 1. The left figure illustrates how categorical features (the dots on the two far left lines) can be combined to form categories in a 2-dimensional space. The right graph shows the development of categorical features in one quality dimension of an agent’s meaning space during the agent’s childhood. The x-axis shows the time in guessing games and the y-axis shows which values are occupied by a categorical feature. The solid lines show the CFs present at a certain time step and the dotted lines indicate the sensitive range of the CFs. Initially there are a few CFs, which are sensitive to a wide area in the feature space. Later on, as more CFs are constructed, the range of the CFs becomes more narrow.

Both agents in a guessing game are presented with a context containing a given number of objects. From these objects, both agents extract perceptual features concerning colour (represented by the red, green and blue components of the RGB colour space) and shape (based on the ratio between the object’s area and the area of its smallest bounding box). The thus resulting four features are then categorised. First, for each object, each feature is categorised using a *categorical feature* (CF), which is a region in one quality dimension represented by a prototypical value. Then all the CFs of an object are combined to form a category (Fig. 1, left), which thus represents a region in a 4-dimensional *conceptual space* (Gärdenfors, 2000).

At the start of each agent’s lifetime, the agent has no CFs in its repertoire. In order to communicate about an object, the agent is forced to distinguish the category of one (or more) object(s) from the other objects’ categories in the context by playing a *discrimination game* (Steels et al., 2002). If categorising an object does not yield a distinctive category, the agent expands its repertoire of CFs by constructing new CFs for which it takes the object’s features as exemplars. This way, the agents gradually constructs a repertoire of categorical features. Initially, these CFs are general and sensitive to a wide area of a quality dimension. Over time, when more CFs are constructed, these CFs become more specific and narrow down their sensitivity (see Fig. 1, right). As a result, when a category is used in a naming event, the reference of the used expression can be overextended. It is this property that is subject of the current investigation. (Note that, even though an agent may have only a few CFs in one dimension, in combination with the CFs acquired in the other dimensions, there are many more possible categories. Since the discrimination game only considers whether different objects in a context are

distinctive or not, these few CFs may be employed successfully for a period of time.)

Once the objects are categorised, the guessing game proceeds. The speaker, who selects one object as the topic, searches its grammar for rules with which it can encode an expression which conveys this topic. The grammar contains simple rewrite rules that are either holistic (e.g.,  $S \rightarrow a\text{word}/a\text{meaning}$ ) or compositional (e.g.,  $S \rightarrow A/m1 \ B/m2$ ). Holistic rules take meanings as categories formed by all 4 dimensions as though they are a single atomic concept. The meanings of compositional rules are formed through a combination of categories from conceptual spaces of lower dimension. If the rule is compositional, the agent will have other rules that rewrite the non-terminals (A and B) to words (e.g.,  $A \rightarrow \text{word}/\text{meaning}$ ). Each rule is given a score, which indicates the effectiveness of the rule based on previous guessing games, and which is adapted according to the outcome of a game. When the speaker finds more than one way to encode an expression, it will select the composition that has the highest combined score. If the speaker fails to encode an expression, it invents a new form either holistically or – in the case that it can encode a part of an expression – in relation to one existing non-terminal.

The encoded expression is then uttered to the hearer, who in turn searches its grammar for ways to decode the expression. Each possible parse results in a possible meaning for the expression. All resulting possible meanings are then filtered such that only those meanings that are consistent with the current context remain. If more than one meaning is left, the hearer selects the one with the highest combined score and the object that belongs to that meaning is then guessed as the speaker's topic. This information is then conveyed back to the speaker (similar to pointing), who verifies whether or not the hearer guessed the right topic. If this is the case, the speaker acknowledges success, otherwise, it will inform the hearer which object was the topic (again similar to pointing).

If the game was successful, the agents increase the scores of the rules that were used, while the scores of competing rules are inhibited (a rule is competing if it could have been used in the same situation). If the hearer guessed the wrong topic, the scores of the rules it used are decreased. In addition, the hearer then adopts the expression with the meaning of the correct topic. If the hearer could not decode the expression, it also adopts the expression with the meaning of the topic.

Adopting an expression is done in one of three ways. First, if the hearer could parse a part of the expression with the intended topic (i.e. a part of the expression maps onto one constituent of an already existing rule of which the meaning matches a part of the topic's meaning), the remaining part of the expression is associated with the remaining part of the meaning. Second, if the first method fails, the hearer will try to chunk (or break up) the expression in two. To achieve this, the hearer searches an instance-base that contains all previously used expression-

meaning pairs, and finds those instances that fit a part of the expression-meaning pair to be learnt. If there are such instances, the heard expression is chunked such that it best fits the data acquired so far. Third, if the expression cannot be chunked, the hearer incorporates the expression holistically and adds its association with the topic's meaning to its grammar unanalysed.

It is important to stress that at the start of each agent's lifetime, its grammar is empty. All linguistic knowledge is thus acquired by playing these guessing games. In the simulations, the guessing game model is integrated with the ILM, such that the population of each iteration consists of a number of adults and a number of children. During an iteration, the population plays a given number of guessing games, after which all adults are removed, the children become adults and new children are introduced. This process repeats for a given number of iterations.

### 3. Overextensions and compositionality

In order to test the hypothesis presented in the introduction, two things need to be shown: (1) overextensions increase the tendency for compositionality to emerge, and (2) overextensions provide an antagonising pressure against compositionality. As mentioned in the previous section, overextensions in this model tend to emerge due to the gradual development of categorical features, as a result of which categories are initially sensitive to a wider range of objects (cf. Fig. 1, right). So, for example, if the agent in Figure 1 learns the word for triangle – for which the proper CF has value 0 – very early in its life, say around guessing game 20, then this word would wrongly be associated with the CF labelled b, which corresponds to a hexagon. Likewise, when this agent needs to produce a word associated with the CF labelled b during the same period, this word could be overextended to nearly all shapes. (Note that in the current study, children only start producing utterances once they are adults.)

To avoid the emergence of overextensions, it is possible to equip each agent with a predefined set of CFs that have a one-to-one correspondence to the features of all objects. In a previous study where this was done, it was shown that compositional structures tend to emerge more rapidly when the agents go through a period of overextensions than when they do not (Vogt, 2005a).<sup>b</sup> As the only difference in these two conditions was the presence or absence of overextensions, this result supports part (1) of the hypothesis.

In the same study, it was shown that when the language is not transmitted through a bottleneck (i.e. all children observe the entire language during their learning period), the compositional structures that arise when there are no overextensions remain stable over subsequent generations. In the case where there were overextensions, the compositional structures tended to collapse in favour of holis-

---

<sup>b</sup>Note that the focus in Vogt (2005a) was not on overextensions, but on statistical properties of the input to language learners.

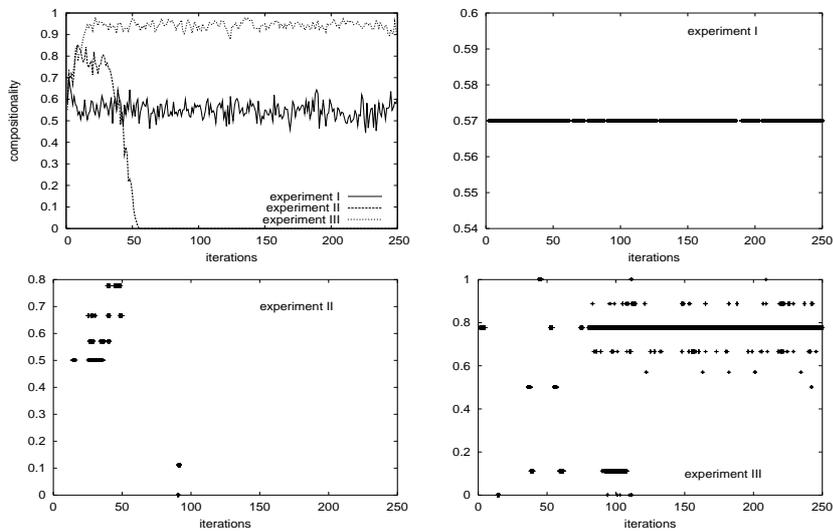


Figure 2. Top left: Compositionality of a typical run of the three experiments measured at the end of each iteration. The other graphs show the typical dynamics of preferred meanings in the shape dimension of a word that was used successfully in experiments I (top right), II (bottom left) and III (bottom right).

tic languages when the language was not transmitted through a bottleneck. This, thus, proves part (2) of the hypothesis.

Now let us take a closer look at some of the dynamics of these findings. In the simulations presented here, the population size was set to 6 (3 adults and 3 learners), the world contained 120 objects (10 colours  $\times$  12 shapes), the population played a total of 3,000 guessing games per iteration, and the simulations were run for 250 iterations.<sup>c</sup> At the end of each iteration, the population was tested in 200 situations, where each agent produced expressions about the same objects and each other agent tried to interpret each produced expression without learning.

From these test phases, the proportion of expressions that were made using compositional rules was measured. Figure 2 (top left) shows the evolution of this *compositionality* measure for typical runs in three different experimental settings (see Vogt, 2005a, for a statistical analysis). In experiment I, the CFs were predefined and the language was transmitted without a bottleneck. As the graph shows, in this experiment, compositionality rapidly emerged to a level around 0.55 and remained stable at this level. In experiment II, the CFs were not predefined and the language was again transmitted with no bottleneck. Here compositionality rapidly increased to a level near 0.8, but collapsed a little later within a few itera-

<sup>c</sup>These are the same parameter settings used in Vogt (2005a, 2005b).

tions to a level of 0. In experiment III, again the CFs were not predefined, but this time the language was transmitted through a bottleneck in which the agents only communicated about 50% of all possible objects. Clearly, when the language was transmitted through a bottleneck, compositionality kept rising until a stable system emerged, with a level of compositionality above 0.9. A similar result is achieved if the CFs are predefined and the language is transmitted through a bottleneck (not shown here, but see, Vogt, 2005a).

The three other graphs in Figure 2 show the typical dynamics of a word that is used to express a certain shape. These graphs were generated by plotting at each 50th guessing game, for each agent, the preferred meaning (i.e. prototypical value of the CF) of a hand-selected word for shape that the population used with a relatively high degree of success. When the CFs are predefined (experiment I), all agents use the same meaning invariably over time. The two other graphs, in which the CFs are not predefined so that these languages are subject to overextensions, show that different agents quite frequently tend to use the same word to express different meanings. For experiment II, the words for expressing shape (or colour for that matter) tend to die when the language becomes holistic, from generation 50 onward. For experiment III, however, there is a clear tendency for the majority of agents to prefer the same meaning for a word.

#### **4. Discussion**

In this paper the effect of overextensions on the emergence of compositionality in language is investigated. The simulations show that when agents are subject to overextending words as a result of the incremental construction of categories, compositionality tends to emerge to a higher degree but the population have more difficulty in arriving at a shared system. The latter is due to the fact that when children hear a particular word early in their development, this word may be associated with a category that has a wide scope. Once the children have acquired all categorical features in a given dimension, this word may then be associated with the wrong meaning. Nevertheless, due to the limited number of interactions among the agents, such associations may survive. This may also happen because other agents may be able to understand such a word in a particular context. As a result, the meaning of a word can drift from one position in the conceptual space to another over time (cf. Fig. 2 bottom left).

When there is no bottleneck on linguistic transmission, the compositional systems that tend to emerge initially all collapse after a period of time. As argued elsewhere, this has to do with the lack of pressure to form compositional structures when the need for them is not there, which is the case if children can learn from the entire language of their predecessors (Kirby et al., 2004; Vogt, 2005b). However, the current study – as well as the one presented in Vogt (2005a) – shows that when there are no overextensions during development, this collapse does not occur. So it is not only the lack of pressure due to the absence of a bottleneck, but

also the additional difficulties in arriving at a shared system due to overextensions that make compositional structures less stable than holistic ones. Intuitively, this can be understood by realising that a meaning drift in one dimension (i.e. linguistic category) of a compositional system affects a larger part of the language than a meaning drift in one dimension of a holistic system (see also, Vogt, 2005b).

Concluding, overextensions arising from the ontogenetic development of meanings in this model do indeed cause both a positive and a negative effect on the emergence of compositional structures. Now, is it possible to extrapolate from this study to the case of human language evolution (and its acquisition)? This question is hard to answer, because the current model is a highly simplified model of human language evolution. Perhaps the easiest part is to find evidence for the positive effect. If the hypothesis is extensible to natural language, the results would predict, e.g., that children – while going through the phase of overextensions – become increasingly proficient at forming compositional structures. Currently, research is underway investigating to what extent such a tendency can be detected in child language acquisition. The negative effect that language becomes more unstable is harder to assess, because this effect occurs only in the absence of a bottleneck, which seems very unlikely for young children learning language (Vogt, in press). However, on the other hand, the model would predict that due to overextensions, differences in preferences on language use would emerge, though, again, this will be very hard to assess empirically.

## References

- Briscoe, E. J. (Ed.). (2002). *Linguistic evolution through language acquisition: formal and computational models*. Cambridge: Cambridge University Press.
- Clark, E. V. (2003). *First language acquisition*. Cambridge: Cambridge University Press.
- Gärdenfors, P. (2000). *Conceptual spaces*. Bradford Books, MIT Press.
- Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals: linguistic adaptation through iterated learning. *Studies in Language*, 28(3), 587-607.
- Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language*. Oxford, UK: Oxford University Press.
- Vogt, P. (2005a). Meaning development versus predefined meanings in language evolution models. In L. Kaelbling & A. Saffiotti (Eds.), *Proceedings of ijcai-05* (pp. 1154–1159). IJCAI.
- Vogt, P. (2005b). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence*, 167(1–2), 206–242.
- Vogt, P. (in press). On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior*, 13(4).