Iterated learning and grounding: from holistic to compositional languages

Paul Vogt
ILK/Computational Linguistics and AI
Tilburg University, Tilburg, The Netherlands.
Language Evolution and Computation Research Unit
University of Edinburgh, UK.
paulv@ling.ed.ac.uk

http://www.ling.ed.ac.uk/˜paulv

Abstract

This paper presents a new computational model for studying the origins and evolution of compositional languages grounded through the interaction between agents and their environment. The model is based on previous work on adaptive grounding of lexicons and the iterated learning model. Although the model is still in a developmental phase, the first results show that a compositional language can emerge in which the structure reflects regularities present in the population's environment.

1 Introduction

Evolutionary computational linguistics has become a booming research area during the past decade, see, e.g., [5, 9] for overviews. One particular area that has gained increasing attention is the emergence of compositional languages, i.e. languages in which parts of an expression have a structured relationship to their semantics. Strikingly, but not surprisingly, almost every study that has investigated the emergence of compositional languages have assumed a predefined meaning space, e.g., [1, 8]. Consequently, these studies are subject to the *symbol grounding problem* [7], which relates to the question how symbols become meaningful to the agent who uses them. One of the reasons why we should try to avoid the symbol grounding problem is that a lot of linguistic structures may be induced from an agent's interaction with their environment.

The only known study that investigates the origins of grammatical structures in a grounded setting has been reported by Steels [10]. In this work, Steels proposes a model in which agents construct a procedural grammar of which the semantics are acquired through their interaction with their environment and the grammatical structures through a complex interplay between the semantics and linguistic utterances produced by the agents. The experimental framework on which Steels' model is based is called the *Talking Heads experiment* [11]. In this experiment, a population of agents attempt to develop a language with which they communicate about aspects of their environment. This environment contains geometrical coloured figures that the robotic agents see with their steerable camera heads.

This paper presents a new computational model based on a simulation of the Talking Heads experiment. This model combines the *iterated learning model* as proposed by Kirby [8] with aspects of symbol grounding aimed at researching the emergence of compositional languages. In particular, the paper attempts to show how learners can induce syntactic and semantic structures by observing the linguistic behaviours of adult speakers and by discovering visual regularities in their environment. Through a process of invention and induction, a language is bootstrapped from scratch and transmitted culturally to subsequent generations.

The following section presents some background on the state-of-the-art in iterated learning and grounding. Section 3 presents the proposed model. Initial results are presented in Section 4, which are discussed in Section 5.

2 Iterated learning and grounding

The iterated learning model (ILM) has been proposed to study aspects of language evolution, in particular the way language is transmitted culturally from one generation to another [4, 8]. The ILM contains a population of adults and learners, where the adults teach their language to the learners through linguistic interactions such as language games. After each iteration, the adults are replaced by the learners and new learners enter the population. Kirby [8] has shown how the ILM could model a transition from initial holistic protolanguage into compositional languages. In this study, the holistic language was constructed from associations between predefined meanings (represented as predicate-argument structures) and unstructured signals. Using a number of heuristics, the agents were able to induce syntactic structures relating to the regularities of predicate-argument structures of the semantic space. By applying a bottleneck on the transmission of the language, Kirby was able to show how compositional languages emerged after a number of generations. One shortcoming of Kirby's simulations is that it was subject to the symbol grounding problem.

In [15], I applied the ILM to study the evolution of lexicons in a simulation of the Talking Heads experiment [11]. In this study, a holistic language emerged based on agents' observations of their environment – thus providing a way of grounding – and on the learners' observations of linguistic behaviours of adults. In these simulations, a shared lexicon was constructed by processing numerous language games in which agents tried to convey the meanings of observed objects. These meanings were adaptively formed using discrimination games. If the agents failed to convey a meaning, new (holistic) word-forms were invented, adopted or associations were weakened. If the agents were successful, used associations were strengthened. In this paper the Talking Heads simulation – implemented in the toolkit *THSim* [16][1] – is used as the starting point for studying the emergence of compositional languages.

The model proposed in this paper is based on finding conceptual spaces. In line with Gärdenfors [6], I will use the term *conceptual space* as a space where concepts (or meanings) can be stored and observations can be conceptualised. A conceptual space is spanned by a number of *quality dimensions*. Each quality dimension relates to some quality (or feature) that can be measured by an agent's sensors. For instance, the qualities Red, Green and Blue are quality dimensions of a conceptual space for colour in artificial agents. For holistic languages, I assume that there is one conceptual space that is spanned by all possible quality dimensions – I call this space the *holistic conceptual space*. In compositional languages, conceptual spaces are of lower dimension and relate to certain qualities such as colour or shape. According to Gärdenfors, a conceptual space can form the semantic representation of linguistic categories [6]. For the current study, I hypothesise that in language, holistic utterances (represented in holistic conceptual spaces) evolved first, and compositional structures emerged from these holistic utterances, cf. [8, 17]. Note that I do not claim that holistic utterances initially referred to holistic

[1] Current version THSim v3.2 can be downloaded from www.ling.ed.ac.uk/˜paulv/thsim.html. This version is still based on holistic signalling only, future releases will include the compositional model described in this paper as well.

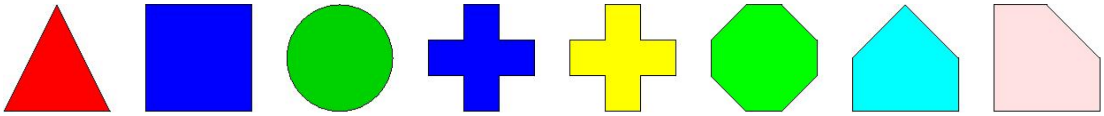


Figure 1: Some of the figures that can occur during a language game.

conceptual spaces, I merely adopt the hypothesis for practical reasons. Given this assumption, to find linguistic categories in a holistic conceptual space, it suffices to discover conceptual spaces of a lower dimension.

Discovering conceptual spaces is guided by a two way process in language development: On one hand, semantic structures are induced from regularities in the interaction between agents and their ecological niche, though constrained by the syntactic structures of their language. On the other hand, syntactic structures are induced from regularities in culturally transmitted linguistic utterances, though constrained by the semantic structures. This principle is based on the findings that language and meanings co-develop [3]. The next section will describe the implemented model in detail. The induction steps in the model are adapted from Kirby's [8] model to integrate symbol grounding.

3 Discovering conceptual spaces

As mentioned, the model is implemented as an extension of the THSim toolkit [16]. In this tool a population of agents can play a series of various language games to develop a language that allows them to communicate about coloured geometrical figures that are displayed on the screen. Whenever the agents fail to communicate successfully, they adapt their ontology (i.e., the set of meanings) and linguistic knowledge (lexicon and grammar) to increase their performance in future games. In the current paper, the population contains only one adult and one learner; the adult takes up the role of speaker, while the learner acts as hearer. After playing a number of language games, the learner replaces the adult and a new learner without any linguistic knowledge enters the population. The period in which the population remains constant is called an *iteration*.

3.1 Sensing the environment

At the start of a language game, a context C of geometrical coloured figures (or objects o_i) is generated by the environment. Each figure is randomly selected from 10 different shapes such as rectangles, circles, triangles, crosses and 6 other regular and irregular polygons (Fig. 1). In addition, each figure is given a colour selected arbitrarily from a set of 12 different colours. So the environment contains a total of 120 objects. In the current presentation, each language game concerns a different context of 8 objects.

The agents 'look' at the context and obtain for each object $o_i \in C$ a feature vector $\mathbf{f}_i$. A feature vector contains a number of features (or qualities) measured by the agent. Currently, the agents use four features, so $\mathbf{f}_i = (f_\mathbf{r}, f_\mathbf{g}, f_\mathbf{b}, f_\mathbf{s})$, where $f_\mathbf{r}$, $f_\mathbf{g}$ and $f_\mathbf{b}$ relate to the $\mathbf{rgb}$ colour space representation of the object, and $f_\mathbf{s}$ is a *shape feature*. The way these features are calculated is not relevant for this paper (see [16] for a description), it suffices to say that they can be measured by a real robot and that each shape has a distinct shape feature. Unlike the real Talking Heads, the features are measured without any noise. This is done to reduce – for the time being – the complexity of the study. After the agents obtained feature vectors for all objects in the context, the context can be described as $C' = \{\mathbf{f}_1, \ldots, \mathbf{f}_N\}$, where $N = 8$ is the context size.

Once the context is set, the speaker of the language game arbitrarily selects one object from the context as the *topic* of the game and informs the hearer which object this is. This strategy of informing the hearer about the reference of the game is based on establishing joint attention and is called the *observational game* [13]. It differs from the *guessing game*, which was originally played in the Talking Heads experiments [11], where the speaker provides corrective feedback after the hearer guessed what the topic was.

3.2 Meaning formation: the discrimination game

Given the context C' and the topic o_t (described by $\mathbf{f}_t$), the agents try to form a meaning to represent the topic. One way to form meanings is to use the discrimination game model, e.g., [11, 13], which is played by an individual agent. The aim of the discrimination game is to find one or more *semantic hypotheses* for a topic that distinguishes the topic from all other objects in the context. Semantic hypotheses are (compositions of) categories that are defined as regions in a conceptual space, represented by a prototype. A prototype is a point in the conceptual space and its category is that region of which all points are nearest to the prototype.

Definition: A conceptual space is said *to cover* certain quality dimensions. The holistic conceptual space covers all quality dimensions, while a non-holistic conceptual space (or *conceptual space* for short) covers only a subset of all quality dimensions.

The way a (holistic) conceptual space is covered is indicated by feature letters. For instance, in the current study there are 4 quality dimensions **r, g, b** and **s**. The holistic conceptual space covers $\mathbf{rgbs}$, the conceptual space for colour covers $\mathbf{rgb}$ and the 'shape space' covers $\mathbf{s}$. If an agent has a holistic category represented by $\mathbf{c} = (1.0, 0.0, 0.0, 0.1)$, this category can be decomposed into two categories covering $\mathbf{rgb}$ and $\mathbf{s}$, represented by $\mathbf{c}' = (1.0, 0.0, 0.0, ?)$ and $\mathbf{c}'' = (?, ?, ?, 0.1)$, where the $?$s are wild cards. If an object is observed by an agent, it could categorise its feature vector holistically (yielding the category set $\{\mathbf{c}\}$) or compositionally (yielding $\{\mathbf{c}', \mathbf{c}''\}$). All categories $\mathbf{c}_i$ of agent a are stored in its ontology $\mathcal{O}_a = \{\mathbf{c}_1, \ldots, \mathbf{c}_p\}$, which is initially empty.

At the start of a discrimination game, the agent categorises all $\mathbf{f}_i \in C'$ by searching those categories in each different conceptual space for which the feature vector is nearest to the category that covers the conceptual space. From these categories, category sets are constructed such that the compositions cover all dimensions of the holistic space. This yields for each feature vector $\mathbf{f}_i \in C'$ a *set of category sets* $\mathcal{C}_{a,i}$.

If all sets are constructed, the agent removes all category sets $H_n \in \mathcal{C}_{a,t}$ for which for some $i \neq t$: $H_n \in \mathcal{C}_{a,i}$, yielding a semantic hypothesis set $\mathcal{H}_{a,t} = \{H_k\}$ for topic f_t. In other words: the semantic hypothesis set contains those category sets for the topic that are not category sets for any other object in the context, thus distinguishing the topic.

If $\mathcal{H}_{a,t} = \emptyset$, the agent has no hypothesis that allows it to distinguish the topic from the rest of the context, and the discrimination game fails. In this case, the agent will create a new holistic category by taking the topic's feature vector $\mathbf{f}_t$ as an exemplar. If $\mathcal{H}_{a,t} \neq \emptyset$, the discrimination game succeeds and the semantic hypothesis set $\mathcal{H}_{a,t}$ is forwarded to the production or interpretation phase of the language game.

3.3 Production

After the speaker has successfully played a discrimination game, it tries to produce an expression to convey the reference to the topic. Production is done in three stages. First, the speaker searches

Grammar	$m \backslash F$	$bluesquare$	red	$triangle$	rue
$r1 = S \rightarrow bluesquare/\mathbf{rgbs}$	$m1 = (0, 0, 1, 1)$	0.6	0.0	0.0	0.0
$r2 = S \rightarrow A/\mathbf{rgb}\ B/\mathbf{s}$	$m2 = (1, 0, 0, 0)$	0.1	0.0	0.0	0.0
$r3 = A \rightarrow red/\mathbf{rgb}$	$m3 = (1, 0, 0, ?)$	0.0	0.5	0.0	0.2
$r4 = B \rightarrow triangle/\mathbf{s}$	$m4 = (?, ?, ?, 0)$	0.0	0.0	0.7	0.0
$r5 = A \rightarrow rue/\mathbf{rgb}$	$m5 = (1, 0, ?, ?)$	0.0	0.0	0.0	0.0
	$m6 = (?, ?, 0, 0)$	0.0	0.0	0.0	0.0

Table 1: An example grammar and lexicon. The left column presents the grammar. The right part of the table shows the lexicon where the forms F are presented in the columns and the meanings m in the rows. The values in the cells represent the association scores $\sigma_{F,m}$.

grammatical rules that fit the semantic compositions of the semantic hypothesis set. Second, the speaker tries to lexicalise the composed categories that fit a grammatical rule. Third, the speaker selects that lexicalisation that has been most effective in the past.

During the agents' lifetimes, each agent a constructs a private grammar $\mathcal{G}_a = \{r_1, \ldots, r_q\}$ with rewrite rules r_i like the ones presented on the left-hand side of Table 1. In this table, the symbol S is the start symbol of a sentence, other upper case letters, such as A and B, are arbitrarily named terminals, the italic lower case strings are word-forms and the bold face strings indicate the covering of the terminals. One might expect that, rather than indicating which conceptual space is covered by the word-forms (as in rules 1, 3, 4 and 5), one could indicate the forms' meanings. However, a form may be associated with more than one meaning (and vice-versa), as shown in the lexicon at the right hand side of Table 1.

Each agent a additionally has a lexicon $\mathcal{L}_a$, defined as an associative memory that associates forms F_i with meaning m_i mediated by an association score σ_{F_i,m_i}. An association score indicates the effectiveness of an element in previous language games. Lexical elements $l_i \in \mathcal{L}_a$ are notated by $l_i = \langle F_i, m_i, \sigma_{F_i,m_i} \rangle$. Initially, both $\mathcal{L}_a = \emptyset$ and $\mathcal{G}_a = \emptyset$.

When searching rules that match the semantic compositions, the speaker searches for a way to parse each semantic hypothesis with the grammar by matching the covers of the composition. Suppose the speaker has obtained the semantic hypothesis set $\mathcal{H}_{s,t} = \{\{m2\}, \{m4, m3\}, \{m5, m6\}\}$. In this case only the first two sets are parseable with respect to the grammar presented in Table 1. The first set $\{m2\}$ fits a rule like $r1$, because it covers $\mathbf{rgbs}$. Likewise, the second set $\{m4, m3\}$ fits rule $r2$ (note that the order of categories is discarded in the semantic hypotheses, the grammatical rules represent the order). The final set $\{m5, m6\}$ does not fit any rule, because the composition covers $\mathbf{rg}$ and $\mathbf{bs}$, which do not combine to form a rule in the grammar.

Given these compositions, the speaker tries to find forms that match the categories of the compositions in the same way as done previously for holistic communication, e.g., [13, 16]. The speaker searches its lexicon for elements of which the meaning matches one of the categories. Continuing our example, composition $\{m2\}$ can be lexicalised with $bluesquare$ and $\{m4, m3\}$ with $triangle$ and red. Thus the speaker has two ways to express the two hypotheses: $bluesquare$ and $redtriangle$, which are derived from compositions $r1$ and $r2 \circ r3 \circ r4$ respectively. Note that the composition $r2 \circ r3$ indicates that that $r3$ is applied to the leftmost free terminal of $r2$. Further note that when an expression is composed of more than one form, the forms are concatenated such that the hearer cannot explicitly detect word-boundaries. Now the speaker will select the expression that was most effectively in the past based on the average association scores of the lexical elements. In the example, the association $\langle m2, bluesquare, 0.1 \rangle$ has an average score of 0.1, while the associations $\langle m3, red, 0.5 \rangle$

and $\langle m4, triangle, 0.7\rangle$ have an average score of 0.6. As the latter is higher, the speaker will express $redtriangle$.

If the speaker fails to produce an utterance, which is the case when it has no grammatical rule to cover the semantics or when it (partially) has no matching association in its lexicon, the speaker expands its grammar and lexicon. There are two possibilities:

1. *The speaker has a rule of more than one constituent that covers the semantics, but there is no matching (or partially matching) association in its lexicon.* In this case the speaker invents one or more new word-forms to associate with the meaning parts for each unassociated category.
2. *The speaker has no rule to cover any of the semantic hypotheses.* In this case the speaker invents a new word-form that is associated with a holistic hypothesis. If no such hypothesis exists, the categories of a compositional hypothesis are merged into a holistic category.[2]

The first case occurs, for example, when the speaker has the semantic hypothesis set $\mathcal{H}_{s,t} = \{\{(1,0,0,?),(?,?,?,1)\}\}$ and the above grammar and lexicon. In that case, it can select rule $r2$, together with rule $r3$ to form a partial expression $red...$. The speaker will then invent a new form, for instance *square*, and adds the association $\langle square, (?,?,?,1), 0.01\rangle$ to its lexicon. In addition, it will add the rule $B \rightarrow square/\mathbf{s}$ to its grammar. (Note that in the simulations forms are invented as sequences of consonant-vowel pairs randomly selected from a finite alphabet.)

The second case occurs, for example, when the speaker has the semantic hypothesis set $\mathcal{H}_{s,t} = \{\{(0,1,0,\frac{1}{2})\}\}$. In that case a new form is invented, say $greenpentagon$, and the association $\langle green\ pentagon, (0,1,0,\frac{1}{2}), 0.01\rangle$ is added to the lexicon. In addition, the rule $S \rightarrow greenpentagon/\mathbf{rgbs}$ is added to the grammar. If the hypothesis set would have been $\mathcal{H}_{s,t} = \{\{(0,1,?,?),(?,?,0,\frac{1}{2})\}\}$, then the two categories are merged into $(0,1,0,\frac{1}{2})$ and the above mentioned adaptations are made. Note that the speaker is not able to invent new compositional structures; it can only exploit existing ones.

3.4 Interpretation and induction

Interpretation Upon receiving the expression, the hearer (or learner) tries to interpret the expression. If it fails, the hearer will try to induce new linguistic knowledge. Interpretation is processed in two stages: parsing the expression and checking the semantics. Parsing is done at the syntactic level, i.e. the expression is parsed relative to the grammar while the semantics is ignored. I will not go into the details of the parser, as this is relatively straightforward. The only complication is that the word-boundaries are not visible. For the time being, the parser results only in one possible parse. In practise, however, there may emerge situations where more than one parse could be possible, but such situations are currently disregarded for practical reasons. The parser results in a list of forms that are interpreted, together with the interpreted composition.

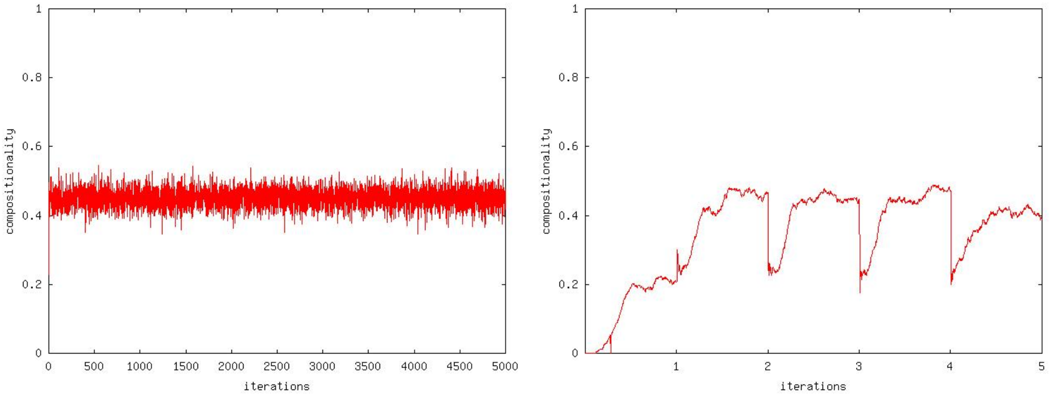
When a parse is found, the resulting list of forms is evaluated relating to the hearer's lexicon and its semantic hypothesis set $\mathcal{H}_{h,t}$ for the topic. So, if the parser returns the result $E = \{e_1, \ldots, e_n\}$, where each e_i is a part of the expression, the hearer searches for each element $e_i \in E$ a lexical element $l_j = \langle e_i, m_j, \sigma_{e_i,m_j}\rangle \in \mathcal{L}_h$ for which the association score $\sigma_{e_i,m_j} > 0$ and $m \in H$, where $H \in \mathcal{H}_{h,t}$ is a hypothesis. A semantic interpretation is complete if the entire expression E can be fully interpreted by a $H \in \mathcal{H}_{h,t}$. If more such interpretations exist, the hearer selects that interpretation H for which the average association score is highest. The language game is successful if the entire expression is completely interpreted by a semantic hypothesis of the topic.

[2] This latter procedure is not ideal, but was implemented to solve impasses occurring when it was not done.

Following our example, suppose the hearer received the expression $redtriangle$. Parsing this expression to the grammar presented above, yields the following expression $E = \{red, triangle\}$ and composition $r2 \circ r3 \circ r4$. If the hearer's hypothesis set $\mathcal{H}_{h,t}$ includes $H = \{m3, m4\}$ or $\{m4, m3\}$, then, given the lexicon of Table 1, this H is the interpretation of $redtriangle$. In this case the language game is a success and the association scores between the used elements are increased by $\sigma = \eta \cdot \sigma + 1 - \eta$, while competing associations are laterally inhibited by $\sigma = \eta \cdot \sigma$. An association is competing if the form matches (part of) the expression but not its meaning or vice versa. Given this scheme, the association scores $\sigma_{red,m3}$ and $\sigma_{red,m4}$ are increased, while $\sigma_{rue,m3}$ is inhibited. The speaker also receives feedback on the outcome and adapts its association scores in a similar way. If the hearer fails to interpret the expression, both agents lower the score of any of the used associations.

Induction When the learner fails to parse the expression syntactically and/or semantically, it will try to induce new linguistic knowledge from the expression with respect to previously learnt knowledge. There are basically three reasons why parsing can fail.

1. *The hearer is able to parse the expression syntactically, but not semantically.* This occurs when the hearer obtained a non-empty expression list E, but failed to find an interpretation. In this case the hearer associates the words of the expression with a $H \in \mathcal{H}_{h,t}$ of equal size and of which the meaning parts cover the conceptual spaces of the terminals of the parsed rule. Going back to our example, suppose the learner received the expression $ruetriangle$, which it can parse using composition $r2 \circ r5 \circ r4$. Further suppose that the only H that covers this composition is $\{(0, 0, 1, ?), (?, ?, ?, \frac{1}{4})\}$. The learner will then add the associations $\langle rue, (0, 0, 1, ?), 0.01\rangle$ and $\langle triangle, (?, ?, ?, \frac{1}{4}), 0.01\rangle$ to its lexicon.
2. *The hearer is able to parse the expression partially on both the syntactic and semantic level.* This happens when the learner finds a composition by which only a part of the expression is interpreted. In this case, the learner associates the not interpreted part of the expression with the remaining elements of the H that is partially interpreted, constrained by the grammar. If there are more ways to interpret the expression partially, the hearer prefers to adapt the remaining part of the expression with the minimum number meaning parts. If there are more than one such partial matches, the one with the highest average association score is selected. For example, if the hearer receives the expression $bluetriangle$ while having a $H = \{(0, 0, 1, ?), (?, ?, ?, 0)\} \in \mathcal{H}_{h,t}$, then, using composition $r2 \circ r? \circ r4$, it is able to match the expression partially with the association $\langle triangle, m4, 0.7\rangle$. The remaining part of the expression $blue$ should then relate to the '$r?$' part of the composition. To achieve this, the association $\langle blue, (0, 0, 1, ?), 0.01\rangle$ is added to the lexicon and the new rule $A \rightarrow blue/\mathbf{rgb}$ is added to the grammar, which then relates to $r?$.
3. *The hearer cannot parse the expression at all.* This occurs when there are no rules in the grammar that match the expression. In this case, the learner tries to split the expression such that it partially matches a split in an existing rule, both syntactically and semantically. Although in principle splits could be applied to every type of rule, they are currently only applied to holistic rules. For example, suppose the learner receives $yellowsquare$ that could relate to $H = \{(1, 1, 0, 1)\}$. A split can then be made in rule $r1$ with the shared form $square$. Additionally, a split can be made in the semantics (if this is not the case, the split is not pursued further), yielding a shared category $(?, ?, ?, 1)$. Rule $r1$ is now rewritten as $S \rightarrow A/\mathbf{rgb}\ square/\mathbf{s}$, and the rules $A \rightarrow blue/\mathbf{rgb}$ and $A \rightarrow yellow/\mathbf{rgb}$ are added to the grammar. In addition the element $\langle bluesquare, m1, 0.6\rangle$ is replaced by the elements $\langle blue, (0, 0, 1, ?), 0.6\rangle$ and $\langle square, (?, ?, ?, 1), 0.6\rangle$. Also the association $\langle (1, 1, 0, ?), yellow, 0.01\rangle$ is added to the lexi-



(a)

(b)

Figure 2: (a) The compositionality during the final 50 games of each iteration. (b) How compositionality evolved during the first 5 iterations.

con.

If no split can be made, the expression is added holistically. This would occur, for example, when the hearer received $greencircle$ or $yellowsquare$ with $H = \{(1,1,0,0)\}$. In this case, a rule with start node S and cover $\mathbf{rgbs}$ is added to the grammar, and the association of the form with a category covering $\mathbf{rgbs}$ is added. If no such category exists, a H composed of two (or more) categories is merged such that the resulting category is holistic again.

3.5 Generalise and merge

When the speaker or hearer has changed a rule, the agent will make sure the grammar contains no redundancies by *generalising* and/or *merging* rules. If two non-holistic rules contain constituents relating to the same linguistic category, these rules can be generalised. For example, if the grammar contains the rules $S \to A/\mathbf{rgb}\ triangle/\mathbf{s}$ and $S \to A/\mathbf{rgb}\ circle/\mathbf{s}$, then both rules are removed from the grammar and replaced by the rules $S \to A/\mathbf{rgb}\ X/\mathbf{s}$, $X \to triangle/\mathbf{s}$ and $X \to circle/\mathbf{s}$, where the terminal X can be any yet unused upper case letter.

If there are two rules that have constituents with different terminals that are acting on the same conceptual space, these rules are merged. For instance, rules $S \to A/\mathbf{rgb}\ B/\mathbf{s}$ and $S \to C/\mathbf{rgb}\ B/\mathbf{s}$ will be merged into $S \to A/\mathbf{rgb}\ B/\mathbf{s}$ and all terminals C are replaced by A throughout the grammar.

4 Results

This section presents the results of a representative simulation. In this simulation, the population contained one adult/speaker and one learner/hearer. The simulation was run for 5000 iterations of 350 language games each. Splitting of utterances was only processed on holistic signals with the consequence that only compositions of two constituents could emerge. The simulation was repeated 10 times with different random seeds.

Figure 2 shows the averaged results of the 10 simulation runs. Graph (a) shows the *compositionality* at the end of each iteration. Compositionality is the average number of compositional expressions

R	iteration 1	iteration 2500	iteration 4999
A	$S \rightarrow x/\mathbf{rgbs}(279)$	$S \rightarrow x/\mathbf{rgbs}\ (110)$ $S \rightarrow A/\mathbf{s}\ B/\mathbf{rgb}(170)$ $S \rightarrow D/\mathbf{gbs}\ gi/\mathbf{r}(48)$ $S \rightarrow B/\mathbf{rgb}\ bi/\mathbf{s}(0)$	$S \rightarrow x/\mathbf{rgbs}(75)$ $S \rightarrow A/\mathbf{s}\ B/\mathbf{rgb}\ (258)$ $S \rightarrow C/\mathbf{s}\ D/\mathbf{rgb}(0)$ $S \rightarrow de/\mathbf{r}\ A/\mathbf{gbs}\ (1)$ $S \rightarrow D/\mathbf{rgb}\ hi/\mathbf{s}(0)$
L	$S \rightarrow x/\mathbf{rgbs}(21)$ $S \rightarrow A/\mathbf{r}\ B/\mathbf{gbs}(5)$ $S \rightarrow C/\mathbf{s}\ D/\mathbf{rgb}(0)$ $S \rightarrow E/\mathbf{rgs}\ ica/\mathbf{b}(0)$ $S \rightarrow F/\mathbf{rbs}\ da/\mathbf{g}(1)$	$S \rightarrow x/\mathbf{rgbs}(1)$ $S \rightarrow A/\mathbf{s}\ B/\mathbf{rgb}\ (45)$ $S \rightarrow C/\mathbf{gbs}\ D/\mathbf{r}\ (1)$	$S \rightarrow x/\mathbf{rgbs}(7)$ $S \rightarrow A/\mathbf{gbs}\ B/\mathbf{r}(3)$ $S \rightarrow C/\mathbf{s}\ D/\mathbf{rgb}(0)$ $S \rightarrow de/\mathbf{r}\ A/\mathbf{gbs}(1)$
A	$S \rightarrow x/\mathbf{rgbs}(156)$ $S \rightarrow A/\mathbf{r}\ B/\mathbf{gbs}(39)$ $S \rightarrow C/\mathbf{s}\ D/\mathbf{rgb}(66)$ $S \rightarrow E/\mathbf{rgs}\ ica/\mathbf{b}(8)$ $S \rightarrow F/\mathbf{rbs}\ da/\mathbf{g}(21)$	$S \rightarrow x/\mathbf{rgbs}(140)$ $S \rightarrow A/\mathbf{s}\ B/\mathbf{rgb}(132)$ $S \rightarrow C/\mathbf{gbs}\ D/\mathbf{r}(44)$	$S \rightarrow x/\mathbf{rgbs}(106)$ $S \rightarrow A/\mathbf{gbs}\ B/\mathbf{r}(32)$ $S \rightarrow C/\mathbf{s}\ D/\mathbf{rgb}(183)$ $S \rightarrow de/\mathbf{r}\ D/\mathbf{gbs}(0)$

Table 2: The grammar of some adults (A) and learners (L) that emerged during one simulation run. Note that the symbol x in the holistic rules is a variable that is filled with different forms. See the text for details.

produced or interpreted by the agents during the previous 50 language games. As the figure shows, the compositionality is already established at approximately 50% from the second iteration onward. Figure 2 (b) shows how the compositionality evolves within the first five iterations. It increases rapidly toward a value near 50% after which it stabilises.

Table 2 shows parts of the private grammars that emerged during one of the simulation runs in iterations 1, 2500 and 4999 of the adult (1st row) and learner (2nd row), and in iterations 2, 2501 and 5000 of the adult (bottom row). The numbers behind the brackets indicate how many times the rules were produced or interpreted during one iteration of 350 language games. At the end of their lifetimes, the agents have around 100 rules, of which about 40 are holistic rules. The table shows all rules that have more than one constituent. Most interesting are the grammars of the adults from iteration ≥ 2. There you see that most of the produced expressions are compositional rather than holistic, although the holistic expressions still make up a great deal of the utterances.

It is also interesting to see that of the compositions made, those that cover the composition $\mathbf{rgb}$ and $\mathbf{s}$ (representing the conceptual spaces for colour and shape) occur most frequently. These reflect the regularities that can be found in the world, where there are a given number of objects (shapes) that are combined with a given number of colours.

Although the table indicates a rather stable grammar (the grammars look very similar), further analysis revealed that the word order flips very frequently (in about 40% of the iterations for rules covering the composition $\mathbf{rgb}$ and $\mathbf{s}$). In addition, this combination is not always the most dominant composition. It competes strongly with other compositions, most notably those covering $\mathbf{r}$ and $\mathbf{gbs}$, which reflect regularities of the $\mathbf{rgb}$ colours used. The colour/shape compositions are inexistent in less than 1% of the iterations, while they are most dominant in about 65% of the iterations.

5 Discussion and conclusion

This paper presents a new computational model to study the origins and evolution of compositional languages of which the semantics are grounded in the population's interaction with their world. The model combines previous work on the emergence of syntax [8] and lexicon grounding [11, 13] with the idea of discovering conceptual spaces [6].

The first results do not show the expected transition from holistic protolanguages to compositional languages, as was obtained by Brighton and Kirby [4, 8]. Their results were obtained by imposing a bottleneck on the transmission of the language from one generation to the next, similar to the poverty of the stimulus. In the current simulation no such bottleneck was imposed (more language games were played than there are objects), because when it was used, compositional languages emerged very infrequently within iterations and died away immediately after. However, looking at the number of meanings that were formed, one could argue that there was a bottleneck, because there emerged roughly an average of 105 meanings, including those covering non-holistic conceptual spaces. Nevertheless, no sudden transition toward compositional languages, after a large number of iterations with holistic languages was observed.

Whether this finding is fundamental or not remains to be seen. It might be caused by a wrong parameter setting, such as the size of the alphabet. This size controls the probability of finding regularities in the initially unstructured expressions, which guides the formation of compositional rules in the splitting part of the inducer. In the simulation presented, the alphabet contained only 6 consonants and 3 vowels, which may be rather small. Future work should investigate the effect of this parameter more carefully.

Another aspect that requires more attention is the selection of rules during production and interpretation. At the moment, the parser stops when it has found a possible parse, which is the first parse that occurs in its grammar. However, it is well possible that an agent has more ways to parse a sentence or semantic hypothesis. If the agent can select a rule from different possible parses, for instance, based on the effectiveness of the rule in previous language games, the transmission of the language may become more stable. Candidates for such selection-based learning algorithms are, e.g., *data-oriented parsing* [2] and *alignment-based learning* [12].

In addition, the discrimination game is likely to be a source for the instability of the emerging language. As the development of the agents is asynchronous, they have a different trajectory of constructing meanings. Furthermore, a discrimination game can succeed even if the semantic hypothesis are no direct representations of the topic's feature vector. This is because the semantic hypotheses are formed from categories that are *nearest* to the topic's feature vector and that distinguish the topic from other objects in the context. Hence, because the context is of limited size and does not include all objects in the world, the categories need not have a prototype that is in a (very) close proximity of the feature vector. An alternative method would be what I have called the *identification game* [14], where a feature vector is categorised with the nearest prototype that is within a certain distance. If no such category exists, a new one is constructed by taking the feature vector as an exemplar. If the threshold distance is sufficiently low, the emerging ontology would correspond more closely to the world. In the current study, the threshold could be set to a value asymptotically approaching 0 and the ontology would resemble the observed objects exactly, because the sensing is not subject to noise. However, this would not be interesting as it makes the grounding trivial, which in reality is not the case.

Nevertheless, the results show that a compositional language which reflects the structure of the world can emerge. The rules most frequently used were composed from colour and shape conceptual spaces, whereas the second most frequently used rules were composed from conceptual spaces that contained a major regularity of the colour space (the **r** component) and the remaining quality dimensions. The mentioned improvement on rule selection could help to favour the most regular aspects of the world. In addition, an incremental statistical analysis of the holistic conceptual space and how it is used could be used to decide how new conceptual spaces should be constructed more reliably.

To conclude, the proposed model for evolving compositional languages grounded in the populations' ecological niche is a promising model to further investigate this problem, although more research is required to improve the model. We are still far from understanding how human language evolved and models such as the one presented here can help to increase our understanding of language evolution. One aspect this model has shown is that languages may be shaped, at least to some extent, by the way in which language users interact with the world.

Acknowledgements

The work in this paper was done as part of a Visiting Research Fellowship at the Language Evolution and Computation unit of the University of Edinburgh, sponsored by the Royal Society of Edinburgh and the Caledonian Science Foundation. The author is very grateful to the members of LEC for their hospitality and support. Special thanks to Henry Brighton and Simon Kirby for their comments on earlier versions of this paper.

References

[1] J. Batali. Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight, editors, *Approaches to the Evolution of Language*, Cambridge, UK, 1998. Cambridge University Press.

[2] R. Bod. *Beyond grammar – An experience-based theory of language*. CSLI Publications, Stanford, CA, 1998.

[3] M. Bowerman and S. C. Levinson, editors. *Language acquisition and conceptual development*. Cambrige University Press, Cambridge, 2001.

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

[5] A. Cangelosi and D. Parisi, editors. *Simulating the Evolution of Language*. Springer, London, 2002.

[6] P. Gärdenfors. *Conceptual Spaces*. Bradford Books, MIT Press, 2000.

[7] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

[8] S. Kirby. 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, 2001.

[9] S. Kirby. Natural language from artificial life. *Artificial Life*, 8(3), 2002.

[10] L. Steels. The emergence of grammar in communicating autonomous robotic agents. In W. Horn, editor, *Proceedings of ECAI-2000*, Amsterdam, 2000. IOS Press.

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

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

[13] P. Vogt. Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication*, 4(1):89–118, 2000.

[14] P. Vogt. Grounding language about actions: Mobile robots playing follow me games. In Meyer, Bertholz, Floreano, Roitblat, and Wilson, editors, *SAB2000 Proceedings Supplement Book*, Honolulu, 2000. International Society for Adaptive Behavior.

[15] P. Vogt. Grounded lexicon formation without explicit meaning transfer: who's talking to who? In *Proceedings of ECAL*. Springer-Verlag, 2003.

[16] P. Vogt. THSim v3.2: The Talking Heads simulation tool. In *Proceedings of ECAL 2003*. Springer-Verlag, 2003.

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