Emergent Functional Grammar for Space
Michael Spranger and Luc Steels

This paper is the author's draft and has now been officially published as:

Abstract
This chapter explores a semantics-oriented approach to the origins of syntactic structure. It reports on experiments whereby speakers introduce hierarchical constructions and grammatical markers to express which conceptualization strategy hearers are supposed to invoke. This grammatical information helps hearers to avoid semantic ambiguity or errors in interpretation. A simulation study is performed for spatial grammar using robotic agents that play language games about objects in their shared world. The chapter uses a reconstruction of a fragment of German spatial language to identify the niche of spatial grammar, and then reports on acquisition and formation experiments in which agents seeded with a ‘pidgin German’ without grammar are made to interact until rudiments of hierarchical structure and grammatical marking emerge.

1. Introduction
The syntactic structure of human languages is based on assigning words to different word classes or parts of speech (such as noun, adjective, adverb, preposition, etc.) and grouping words into phrases (such as a nominal phrase), which can then be combined into larger phrases. From the viewpoint of functional grammar, word classes suggest possible grammatical functions and phrases combine these grammatical functions into larger units which in turn have their own grammatical functions (Dik, 1978; Halliday, 1978; Hengeveld & Mackenzie, 2008). This chapter reports on preliminary investigations on how these structures could originate.
Two approaches have been discussed for the emergence of grammar in the literature: a syntax-oriented and a semantics-oriented approach. Syntax-directed approaches are in line with theories of language processing that favor exemplar-based inventories (Daelemans & Van den Bosch, 2005) and with usage-based models of language learning (Barlow & Kemmer, 2000). Recurrent combinations of words, so called collocations or, in the case of structures, collocations (Stefanowitsch & Gries, 2003), are assumed to be stored as such in memory and thus become available as patterns that can be re-used. These patterns are then generalized, yielding templates for subsequent patterning. Agent-based simulations in this direction have been reported in (Hashimoto & Ikegami, 1996; Vogt, 2005; Gong, 2011). Syntactic categories like nouns or adjectives and hierarchical structures appear as a side effect of inductive learning and the notion of a grammatical functions is absent.

Semantics-oriented approaches hypothesize that the main source of hierarchical structure in language comes from the fact that semantics is compositional. For example, the words in the nominal phrase “the big red block” form a group because the meanings of the individual words each contribute towards the communicative goal of the phrase as a whole, namely to identify an object in the present context. Parts of speech are in this approach signalling possible grammatical functions that words or phrases can have. For example, nouns usually introduce a class of objects which can be used to pick out a set of possible referents from the context, adjectives introduce predicates that can be used to further narrow down this set, articles indicate how to pick out the referent or set of referents of the phrase as a whole. Phrases signal which conceptualization strategy is intended. A conceptualization strategy combines a number of semantic functions into a more encompassing network which achieves a specific communicative goal, such as object reference. For example, the nominal phrase “the block on your left” evokes a strategy where a particular object (“the block”) is identified through a spatial relation (“left”) which is applied from the perspective of the listener (“your”). Earlier chapters, and particularly the previous chapter by van Trijp (2012) already demonstrated the power of a semantics-oriented approach to explain the origins of case grammar. The same perspective is now utilized to explore the origins of hierarchical structure, grammatical functions, and word classes.

The domain of spatial language will be our test bed. This domain is appropriate because a large amount of typological data is available for spatial language and important literature exists on spatial cognition. Moreover spatial grammar shows as much cross-linguistic variation as spatial lexicons (Levinson & Wilkins, 2006), suggesting that cultural processes play a crucial role in forming and coordinating
the spatial grammar used in a particular language. By using spatial language as our test domain, we can also build on an earlier chapter in this book (Spranger, 2012) which showed how spatial terms and spatial categories can emerge in a distributed population of agents playing situated embodied language games.

We follow the same methodology as employed in earlier chapters, namely, conduct reconstruction experiments to get empirical grounding in real language phenomena, conduct acquisition experiments showing that the language strategies being proposed are adequate to learn a realistic grammar, and then conduct grammar formation experiments to show how a new grammatical system may come off the ground through a process of cultural evolution. German spatial language has been used for the reconstruction experiments and the grammatical constructions necessary for expressing spatial relations have been formalized and tested in Fluid Construction Grammar. This grammar is discussed at length in Spranger & Loetzsch (2011). Here we can therefore focus immediately on the acquisition and emergence of spatial language. But first the role expected to be played by grammar must be clarified and this requires in turn that the approach to spatial semantics to be used in the experiments is briefly discussed.

2. The role of grammar

2.1. Spatial semantics

The experiments in this book start from the hypothesis that the primary function of language, and hence the most likely stimulus for its origins, is to support embodied interactions about objects and activities in the real world. It follows that the semantics needed for language has to be grounded in the world through the sensorimotor system of speakers and listeners. When the speaker says “the block left of the box” the hearer has to pick out the block and the boxes in the shared context and then use the spatial relation left-of to identify which block is intended. The hearer therefore has to perform a set of operations (further called cognitive operations) over his private internal model of the situation enhanced with virtual objects and properties constructed during semantic processing. The speaker has to make a plan which operations the hearer should perform, so that his communicative goal is achieved.

To properly model and test this approach requires experiments with physical robots that are also grounded in the world through a sensorimotor system and have a body which they can use as a basis for structuring the world. The experiments reported here have all been done using the robotic set-up illustrated in figure 1.
Figure 1. This figure shows the set-up for spatial language game experiments. It involves two robots, a set of blocks, a box with tags indicating front, back or left and right side, and a global landmark. Left and right show the internal world model as perceived by the left and right robot. Grounded language semantics is achieved by cognitive operations over these internal models.
Our experiments in cultural language evolution use a representational system called IRL (Incremental Recruitment Language) to represent the cognitive operations and the entities over which they operate (Steels, 2000; Van den Broeck, 2008; Spranger et al., 2010, 2012). This system builds further on Intensional Logic based approaches, such as Montague grammar (Montague, 1973; Partee, 2003), and procedural semantics approaches to computational linguistics (Woods, 1986; Johnson-Laird, 1977).

The meaning of an utterance is a mental program that combines a set of semantic functions operationalized as cognitive operations into a network. An example is shown in Figure 2 for the utterance “links des blockes” (to the left of the block):

(1) \text{links} \quad \text{des} \quad \text{Blockes}
\text{left.PREP.GEN} \quad \text{the.DET.GEN} \quad \text{block.GEN}

‘to the left of the block’,

Each node in the network invokes a particular semantic function, such as bind, construct-region-lateral, apply-selector, etc. These functions have a number of arguments. The arguments are specified using variables which are denoted as labels with a question-mark, such as \(?\text{region}\), \(?\text{landmarks}\), \(?\text{cat}\) (for category), \(?f-o-r\) (for frame of reference), etc. Semantic functions are connected to each other when they share the same variable, and these connections are drawn as lines in the graphical representation of the network. One can think of these variables as slots that can be filled, or as we say ‘get bound’, to specific values. For example, the variable \(?\text{cat}\) can be bound to a particular category labeled left.

Note that the names of the semantic functions and the labels of the variables are just conventions. Whether we call a function apply-selector or rotceles-yllppa or a variable \(?\text{region}\) or \(?\text{noiger}\) does not make any difference in the operation of the system. The only meaning all these terms have is determined by the operations defined over them. We just use labels that make sense in order to make the networks more comprehensible for human readers. Given that typical experiments involve thousands of such names, this is not a luxury.

Here are examples of some of the semantic functions used in this network:

- The bind operation binds a value to a variable. The variable is typed which is relevant for transforming meanings to language or in retrieving stored mental programs based on partial programs. For example,

  \text{(bind lateral-category \(?\text{cat}\) \text{left})}

  will bind the category labeled \text{left} to the variable labeled \(?\text{cat}\). The type of this variable is \text{lateral-category}.
The `get-context` function binds its argument to the set of objects in the present context. For example if the context contains two objects, obj-265, obj-266, then after executing the following operation (bottom left of the network in Figure 2):

\[ \text{(get-context ?ctx)} \]

The variable ?ctx becomes bound to the set \{obj-265, obj-266\}.

The function `apply-class` has three arguments: a target set of objects, a source set of objects and some object-class that can be used to filter the set of source objects to determine the target objects. It is used in the example network in Figure 2 in this way:

\[ \text{(apply-class ?landmarks ?ctx ?landmark-class)} \]

The target set is here ?landmarks, the source set ?ctx and the class ?landmark-class. The variable ?landmark-class gets bound to the object-class block in another bind operation and `apply-class` can then filter the objects in the context (bound by the `get-context` operation to ?ctx) to retain those that satisfy the prototype of a block.

The function `construct-region-lateral` (used near the top of the networks shown in Figure 2) has five arguments. It takes a set of objects (here equal to ?ctx) and uses a landmark (bound to ?landmark) and a category (bound to left by another bind operation) to construct a virtual region (bound to ?region) given a particular frame of reference (bound to ?f-o-r). This lateral region is an example of new virtual entities that are constructed on the fly as a network is executed.

IRL comes with a library of hundreds of semantic functions which can be recruited by agents to assemble such networks. Most of these functions (such as `apply-class` or `get-context`) are entirely generic.

Other sentences will have other mental programs as their semantics. The network for following Example 2 is shown in Figure 3.

\[ (2) \quad \text{der} \quad \text{linke} \quad \text{Block} \]
\[ \text{the.NOM left.ADJ.NOM block.NOM} \]

‘The left block’,

And the mental program for Example 3 is shown in Figure 4.
Figure 2. Proposed meaning for the utterance “links des Blockes” (left of the block) This semantic structure encodes how to construct a lateral region based on the spatial relation left using a landmark.

Figure 2

(3) der Block links the.DET.NOM block.NOM left.ADV

‘the block left’,

The IRL system not only takes care of interpreting a mental program in a given sensory context. It is also capable to plan such a program given a particular goal, a particular context, and an inventory of recruitable semantic functions. And it can complete partial programs to find the one that gives the most plausible result in a particular context. This is particularly important, as we will see, in language acquisition and helps also to make language understanding more robust.

2.2. The Role of Grammar

The power of human language comes to a large extent from the fact that it is an inferential coding system (Sperber & Wilson, 1986). The speaker says as little as possible and the hearer has the daunting task to reconstruct the mental program intended by the hearer and to execute it within the context of his own private situation model, derived by his own sensorimotor system. Speakers certainly do not express anything like the mental programs shown here. Instead, they give hints about which program is appropriate. The first source of information is given by the
Figure 3. Proposed meaning for the utterance “der linke Block”. The semantic structure specifies that a set of blocks is computed (apply-class) which is further filtered by the spatial category left (apply-angular-spatial-category-group-based). The spatial category is applied to a group-based relative landmark (Tenbrink & Moratz, 2003). Finally, the selector unique is applied for retrieving the referent of the phrase.

Figure 4. Semantic structure underlying Example “der Block links” (the block left). The spatial relation left is used to construct a region based on an unspecified landmark. This region is used to filter the context, and the resulting elements are tested to see whether the class block applies to them. And this is followed by an operation to select the unique member of the remaining set.
lexicon, as words provide initial bindings of some of the variables. For example, the words in the utterance “links des blockes” (Figure 5) signal that the mental program will make use of a spatial category (“left”), a selector (“the”), and an object class (“block”). But the other utterances (Examples 2 and 3) use the same ingredients and so the information provided by the lexicon is not enough to uniquely determine which mental program is intended.

This is where grammar is needed. The role of grammar is to provide clues on what conceptualization strategies the listener should use. It does this in two ways. Each lexical item is assigned a particular word class (for example adjective or adverb) which determines the possible grammatical functions of the item. For example, adjectives typically have an adjectival function but they may also have a predicative function or may even be coerced into a nominal function. The grammatical functions then constrain the possible conceptualization strategies that the meanings of the lexical items may play a role in. Lexical items with specific grammatical functions are combinable into phrases which contain additional syntactic markings, such as word order, agreement, morphological markers or intonation contours.

Phrases come in many different types and each phrase provides constraints on what possible conceptualization strategy can be used to interpret the phrase.

- In Example 2, ”der linke block” (the left block), the word “linke” (left) introduces a spatial relation \texttt{left} and signals that it is used in an adjectival function, i.e. to further restrict the set of objects denoted by the noun (Eschenbach, 2004). The nominal phrase as a whole invokes the network shown in Figure 3.

- In Example 5, ”links des Blockes” (left of the block), the word “links” (left) is used with a prepositional function and the same relation is now applied to a landmark denoted by the noun (Wunderlich & Herweg, 1991). This type of prepositional nominal phrase invokes the network as shown in Figure 2.

- In Example 3, “der block links” (the block left), “links” (left) is used in an adverbial function, which can be seen by the lack of a subsequent noun phrase (Eschenbach et al., 2002). The use of a spatial category in an adverbial function entails inferences about the type of the region, e.g., the region can be internal or external, and about the presence of a covert landmark (see (Tenbrink, 2007) for more detailed discussions).
So these three examples show clearly how a grammatical system provides valuable indications as to which network of semantic functions should be used to interpret an utterance.

The relation between parts of speech and conceptualization strategies is not straightforward. The suffix “-ly”, as in “slowly” marks this word as an adverb, but often there is no specific marking and hence the possible parts of speech of a word need to be stored in the lexicon. Moreover there is always the potential for coercion, meaning that a word typically used in one grammatical function can sometimes be used in another function. For example, many nouns or adjectives in English can be used as verbs (as in “I bike home” or “I slow down”) in which case syntactic context and morphological marking provides clues that a coercion has taken place.

Each part of speech is associated with possible grammatical functions (e.g. an adjective like “blue” can function as an adjectival modifier in a nominal phrase as in “the blue box” but also as the predicate as in “the box is blue”), and which grammatical function is actually chosen depends on the syntactic context in which the word occurs. Moreover a single word typically provides information about only one small part of a semantic network, mostly which predicates are involved. Which network is intended depends therefore on the type of phrase in which the word occurs. For example, the word “links” gets a prepositional function in example 5 because it occurs in a prepositional phase which as a whole evokes a network as shown in Figure 2. The nominal phrase which is part of the same phrase (in this case “des blockes”) is providing the landmark that will be used to construct the lateral region.

3. Establishing the niche for grammar

As mentioned earlier, we have already conducted reconstruction experiments for German spatial cognition and grammar (Spranger, 2011a; Spranger & Loetzsch, 2011). These experiments used spatial language games within the experimental set-up shown in Figure 1. This grammar as well as the conceptualization strategies they invoke can now be used to do experiments to examine the proposed role of grammar. Specifically we can do experiments that compare the performance of agents that are scaffolded with a spatial lexicon but no grammar, and therefore use a kind of ‘pidgin German’, to agents with the fully reconstructed spatial grammar of German. A possible phrase in pidgin German is:

(4) link block der
    left block the
The words occur in a random order and there is no syntactic indication of parts of speech nor phrasal structure. The three mental programs discussed earlier (shown in Figures 2, 3 and 4) are all possible ways to interpret this utterance.

There are two important differences between these semantic structures. The first is related to the cognitive operations involved. For instance, the semantic structure in Figure 3 involves the operation apply-spatial-category-group-based which is not found in the ones in Figures 2 or 4.

The second type of difference is related to how cognitive operations are linked, as can be seen when comparing structure 2 and 4. In Figure 2 the landmark of the spatial region is given by the subnetwork consisting of apply-selector and apply-class, whereas for the structure in Figure 4 this subnetwork is linked to the output of the operation apply-spatial-region. The linking expresses the difference in the use of the “der block” (the block). In the structure in Figure 4 the spatial region modifies the set of blocks. In the structure in Figure 2 the block is the landmark of the region.

Summarizing, grammar expresses which cognitive operations are part of the semantic structure of an utterance and how the operations are internally linked. Hence, when grammar is removed, semantic ambiguity increases. Semantic ambiguity is defined here as the number of different possible semantic structures underlying the same utterance (Spranger & Loetzsch, 2011).

The increase in possible interpretations of a phrase impacts communication in two ways.

- First, it can lead to failure in communication because the hearer interprets the phrase differently and the different interpretation leads to mistakes in establishing reference, i.e. the hearer interprets the phrase to refer to the wrong object.

- Second, semantic ambiguity leads to additional effort in interpretation on the part of the hearer. If there are multiple interpretations, of a phrase the hearer has to test all of them, in order, to find the correct interpretation.

Figure 5 shows the impact of grammar on communicative success in different spatial settings, using the spatial setups and language games described in (Spranger, 2012). We compare two populations of agents. One is equipped with the full German locative grammar. The second population only is given lexical constructions and therefore producing only pidgin German.

When an agent equipped with a purely lexical system speaks, the agent only conveys the semantic entities, e.g. spatial relations, object classes, determiners and
Figure 5. The figure compares the performance of agents equipped with a purely lexical system (left column) with a population in which all agents operate a reconstructed German space grammar (right column) for three different conditions (many objects, no allocentric landmark and combined). The role grammar plays increases as environments become more complex.

discourse roles without explicitly marking their roles or relationships in the semantic structure. The figure compares the performance on different sets of spatial scenes with varying features and degrees of complexity. To the left the condition is one in which many objects are distributed around an allocentric landmark (the box). The
middle condition is one in which scenes have no allocentric landmark. The condition shown to the right has both scenes with allocentric and without allocentric landmark but is generally more complex with respect to number of objects and distribution of objects. Additionally, the position of the interlocutors is more varied.

Clearly, the more complex spatial scenes are the more important non-ambiguous communication is. Communication does not break down completely when agents are lacking grammatical devices in their language. Rather, agents are in general successful in communication. They are able to establish correct reference in well over 70% of cases and, for instance in the middle condition, in almost 90% of the cases.

This is due to the powerful active interpretation capacity that IRL provides, which allows agents to interpret utterances robustly even in the face of semantic ambiguity. But, this success is also in part due to the overall limited nature of conceptualization strategies agents are given. For every scene about 10 but not an infinitely large number of conceptualizations of reality are possible. This allows agents to guess the meaning of purely lexical utterances in many scenes. We can conclude that there is a communicative advantage for having grammatical constructions that allow agents to recover additional information not communicated by lexical items alone.

The next question is whether grammar has an effect on cognitive effort, i.e. on processing efficiency. Figure 6 shows the advantage of grammar with respect to processing of spatial utterances by comparing the number of interpretations hearers had to try, in order, to arrive at the best interpretation of the phrase. Clearly this number is significantly higher for lexical systems than it is for the German locative grammar discussed here. For purely lexical systems, the number of interpretations the hearer has to try for each phrase is high. On average more than six interpretations have to be tried for every single utterance in all three conditions. When grammar is used, the number drops to around 1.0, which means that in many cases an agent can directly interpret the utterance without having to try multiple interpretations. This shows the clear advantage for having grammar from a processing point of view.

If grammar has positive effects on processing and communicative success, one can ask the follow-up question: What are the factors determining just how much impact grammar has? How much agents with grammar perform better in terms of communicative success is largely a function of the environment and the number of possible interpretations of each phrase?

One hypothesis could be that the more complex the environment and the larger the number of different possible interpretations of a phrase, the more problems in communication. If agents share similar viewpoints on the scene and if there are
Figure 6. This figure compares the number of semantic structures tried in interpretation for German locative phrases processed with and without grammar for three different conditions. This measures semantic ambiguity and shows efficiency in interpreting utterances. The figure shows the average number of interpretations (over 10000 interactions). In case a grammar is available the average is just barely above 1.0. We conclude that grammar greatly reduces the ambiguity in interpretation.

only few objects in the scene, the effect of grammar is less strong than in cases where viewpoints are different and the number of objects is high (a fact that is demonstrated in Figure 5).

But the relation between scene complexity, semantic ambiguity and communicative success is subtle. For instance, the results shown in Figures 6 and 7 suggest that it is not only the number of interpretations that make a difference, but the ambiguity has to matter with respect to the environment. Figure 7 shows how often the meaning hearers recover from an observed utterance is equal to the meaning the speaker used in conceptualization. The interesting condition is the middle condition which shows the largest number of average interpretations of utterances (Figure 6) but the smallest drop in interpretation correctness. In other words, just because the
Figure 7. This figure compares the semantic structure recovered by the hearer with the conceptualization strategy originally chosen by the speaker. The effect is compared for different sets of spatial scenes. If the semantic structures are equal the interaction counts as 1.0, if not as 0.0. The results show the average over 10000 interactions. In the case of grammatical systems, the speaker was able to recover the correct semantic structure in all games. For purely lexical systems this number drops to 80%. This number correlates to some extent with communicative success. But, not in all spatial scenes is a drop in the number of scenes in which the hearer correctly interprets the phrase equal to to a drop in communicative success.

The next question is what advantage grammar has over purely lexical systems in terms of processing efficiency. The performance with respect to processing is governed by the number of possible different interpretations of a particular phrase. The higher the number of possible interpretations the more conceptualization strategies
need to be tested and processed. The number is essentially a function of how much re-use of lexical items occurs in the language or just how much particular semantic entities such as spatial relations participate in different interpretations. For instance, in the German locative system, lateral and frontal projective spatial relations occur in relative and intrinsic conceptualization strategies. Absolute categories do not participate in intrinsic and relative conceptualization strategies but only in absolute ones. To remove the disambiguating power of grammar, therefore, has less effect on absolute spatial relations (in that respect). Consequently, if a semantic entity only participates in a single conceptualization strategy removing parts of grammar related to that entity has little or no effect, whereas when the entity participates in many different conceptualization strategies this can have a big impact. Additionally, the increase in ambiguity is paired with features of the environment. If features of the environment are such that agents are not using a particularly ambiguous set of strategies then the ambiguity does not play a role in these conditions.

4. Emergence of Grammatical Markers

We have compared in the previous section the role of grammatical markers by cutting away parts of a German grammar and investigating their effect on cognitive effort and semantic ambiguity. The next step is to come up with a language strategy that shows how such a grammar may emerge. Obviously to do this for the full complexity of a system like we find in German grammar is too big a step. Instead we are going to perform a more focused small scale experiment that nevertheless shows the essence and viability of the approach.

We will use the setup discussed earlier and shown in Figure 1. In the initial conditions, we scaffold agents with four conceptualization strategies: intrinsic (illustrated in 8), absolute, relative from the perspective of the speaker and relative from the perspective of the hearer. (See Tenbrink, 2007 and Levinson, 2003 for the difference in intrinsic, relative and absolute conceptualizations). Agents are scaffolded with four angular spatial relations, which are modeled respectively after the German “vor” (front), “hinter” (back), “links” (left) and “rechts” (right) spatial categories. The earlier chapter on spatial terms and categories (Spranger, 2012) showed how such a lexicon might emerge.

Agents are scaffolded with lexical constructions so that they can produce and parse three word utterances consisting of a spatial relation, an article and a noun, as in “links des blockes”. Word order is not relevant. Each lexical item introduces a building block of a possible conceptualization strategy, most importantly a bind operation (see the appendix for an example of a lexical construction). There are no
given parts of speech distinctions (so it does not make sense to talk about nouns or articles), but each word has various associated semantic categories which can act as selection restrictions. For example, the word for left is defined as a spatial category and the word block as an object-class. Because the three words of an utterance alone never distinguish which conceptualization strategy is appropriate, the hearer always has to try all strategies in order to retrieve the one that is most likely in the present context. All four strategies can use the same set spatial relations and all apply the frame of reference to an allocentric landmark.

In the initial conditions the agents have a purely lexical language system and can only use semantic criteria to see how the meanings of the different words may fit together with respect to a specific conceptualization strategy. What we would like the experiment to show is the formation of grammatical constructions that (i) pull together individual words to form a complete phrase (based on semantic criteria only), and (ii) syntactically mark this phrase as evoking a particular conceptualization strategy.

The agents will use self-invented particles as syntactic markers, such as “bo” which might mean to use the conceptualization strategy shown in Figure 2. A grammatical utterance could therefore be “block linke bo der”. Word order still does not play any role at all. So this utterance is equivalent to “linke bo block der”.

The remainder of this section introduces the language strategy that agents are going to use to invent and coordinate such a grammar and then the experimental results testing the adequacy of this strategy.
4.1. A Grammatical Language Strategy

As we have seen in earlier chapters, a language strategy contains three components: a method the speaker can use to invent new linguistic material when it is needed, a method the hearer can use to acquire novel linguistic material, and an alignment process whereby speakers and hearers can coordinate whatever they have invented or picked up from others. As we have seen in all other chapters of this volume, the invention and acquisition methods can be greatly enhanced when agents monitor their own performance, diagnose problems and trigger repairs for them.

Let us first consider invention. The speaker produces an utterance by first finding a conceptualization that would allow the hearer to localize the topic he has chosen. He then uses his lexicon and grammar to come up with an utterance. However before rendering this utterance for the hearer, he re-enters the utterance back in his own parsing and interpretation process, thus simulating how well the hearer might understand it. This way, the speaker can detect whether a problem occurs and repair this by creating a new construction so that this combination of lexical units is no longer semantically ambiguous. This leads to the first repair action:

* Speaker resolves semantic ambiguity

- Diagnostic: The speaker detects through re-entrance that there is more than one possible semantic structure.

- Repair: The speaker invents a new grammatical construction by grouping all the lexical units in a new hierarchical unit, by inventing a new marker (which will be a random string), and by adding this marker to the syntactic constraints of the construction. The meaning associated with the construction evokes the conceptualization strategy that was used.

An example of a construction using the Fluid Construction Grammar formalism is shown in the appendix.

Suppose that the speaker was able to produce an utterance which according to him would lead to a successful interaction. Now the hearer is parsing the utterance and there is the possibility that he has never encountered the marker before. This brings us to the second repair action:

* Hearer acquires a new marker

- Diagnostic: The hearer encounters a string in the input utterance that is unknown.
• Repair: The hearer first tries to guess the possible meaning of the utterance, either because it is clear from the context and the information provided by the lexical items what topic was chosen by the speaker, or by signalling incomprehension in which case the speaker points to the topic he had in mind. On that basis the hearer can himself go through the phase of conceptualizing reality to find back the possible mental program that the hearer could have used. The hearer can then reconstruct the construction that the speaker employed and build a similar construction by combining all lexical components into a new unit and hypothesizing that the unknown marker signaled which conceptualization was intended.

There is a further possibility that the language game fails because the marker used by the speaker is interpreted to mark another semantic structure for the hearer. 

**Hearer misunderstood marker**

• Diagnostic: The game failed because the hearer pointed to the wrong topic.

• Repair: The hearer signals incomprehension and the speaker points to the topic he had in mind. On that basis the hearer conceptualizes reality to reconstruct a mental program that would work for this topic, then builds a new construction by combining all lexical components into a new unit and hypothesizing that the marker used by the speaker signals the alternative conceptualization.

Next to invention and adoption, agents need an alignment strategy which has to achieve two objectives:

1. Variation across the markers for the same conceptualization must be damped so that coherence arises, i.e. that all agents in the population use the same grammatical marker for the same conceptualization. This is similar to the damping of synonymy in the lexicon.

2. Competing conceptualizations for the same marker must also be damped. This is similar to the damping of meaning uncertainty and homonymy in lexical language evolution.

To achieve grammatical coherence, we use the same lateral inhibition strategy as used in earlier papers (see Steels & Loetzsch, 2012 and Spranger, 2012, all this volume), but now operating on the level of grammatical constructions. This means that when a game was successful, the score of the grammatical construction that was
used is increased and that of competing constructions decreased, both by the speaker and the hearer. A competing construction means, for the speaker, another construction covering the same meaning but using another marker, and for the hearer, another construction with the same marker but covering another meaning.

4.2. Experimental Results

We have tested the performance of the proposed invention, adoption and alignment operators on three different sets of spatial conditions: One condition where agents are looking at the scene from the same perspective, a second one where they look at the scene from different perspectives and can possibly use an allocentric landmark, and a third one where there are many more objects.

Figure 9 shows the semiotic dynamics for the first condition. Agents are clearly able to develop a rudimentary spatial grammar. In the end stage four hierarchical constructions with respective markers are used, one for each of the four conceptualization strategies. A typical utterance looks as follows:

(5)  
\[ \text{wekato box the bo} \]
\[ \text{front box the bo.MARKER} \]
\[ \text{‘to the front of the box’} \]

The marker “bo” marks the intrinsic conceptualization strategy. “wekato” can be roughly translated as an intrinsic projective category similar to “in front of”.

Figure 10 shows a similar result for the second condition, namely where agents see the scene from a different perspective. And Figure 11 shows results for the third condition, i.e. where there is a mixture of same and different perspective and also more objects in the scene.

These results show that agents can develop a successful system for disambiguating the semantics of particular phrases. How many constructions emerge depends on three factors 1) ecological conditions, 2) eagerness of invention and 3) number of ambiguous conceptualization strategies.

1) For instance, if absolute features never coincide with intrinsic features, there is no need to disambiguate the two. Consequently, agents do not develop markers to disambiguate these two phrases as no situations arise where that would be required. In the experiment, there is always some reference objects in the scene, namely the robots themselves. All conceptualization strategies use a reference object, so as soon as agents start using a relative conceptualization strategy they need to disambiguate.
Figure 9. Emergence of grammatical constructions in an experiment where speaker and hearer see the scene from the same perspective. The number of possible interpretations (semantic ambiguity) drops as the set of grammatical constructions increases, implying that cognitive effort decreases. The interpretation reconstructed by the hearer from the input utterance also becomes similar to that of the speaker.

Figure 10. Emergence of grammatical constructions where the scene appears from a different perspective.
Figure 11. Emergence of grammatical constructions where the scene appears both from the same or from a different perspective and where many more objects appear so that the communicative task is more challenging.

2) A second influence on the number of constructions in the grammars of the agents is related to how eager agents invent. For instance, if agents only invent when there is the danger that the hearer gets the wrong object, then in cases where the environment is very simple or objects are easily discriminated, agents will develop fewer constructions.

3) A third influence relates to how many conceptualization strategies are available to agents. In the experiments discussed here, four strategies were provided. However, there can certainly be many other spatial strategies and not all of them are necessarily ambiguous. In the case, where agents use only non-ambiguous strategies, the agents will not develop constructions.

5. Conclusions

This chapter explored a semantics-oriented approach towards the origins of functional grammar. Functional grammar provides hints to hearers about the conceptualizations underlying an utterance and the semantic functions of the different concepts involved. We used spatial language as the test domain, starting from reconstructions of German spatial language. A first experiment showed that agents
who can express information about the nature of the conceptualizations underlying a sentence need less cognitive effort and minimize possible misunderstandings, compared to agents which can only use a pidgin German without grammar. A second experiment showed that agents endowed with a strategy to build grammatical constructions are able to self-organize and coordinate an emergent grammar and this grammar effectively decreases cognitive effort and increases communicative success.

The results of this chapter are only a first step in a number of ways. At this moment the constraints which form part of the newly formed constructions are semantic - apart from the extra marker, such as “bo”. So there are no true parts of speech yet because the constraints on the constituents of a phrase are semantic in nature. An obvious next experiment is therefore that agents introduce syntactic categories and associate them with the lexical items that can fill slots in a construction. A further experiment should then focus on the emergence of grammatical functions as an additional layer of description on top of syntactic categories, so that phenomena such as coercion can be handled. This is already the case in the reconstructed German spatial grammar.

Another challenge now is scale-up. The strategies being used here by speakers to build new grammatical constructions and by hearers to acquire a new construction are entirely general and they could be used for any semantic domain, not just spatial language. A next step is therefore to scale up the experiment to conditions where agents have many more possible conceptualization strategies, so that they get to a much richer grammar. In the current experiment, agents are given four strategies, but clearly many more are possible and we have already operational experiments showing how such strategies may arise, see Spranger, 2011a and Spranger, 2011b. Bigger inventories should also provide opportunities for the build up and negotiation of more abstract constructions.

These experimental challenges are very complicated to concretize, particularly in experiments with robotic agents, but we believe the work reported here gives a solid foundation on which to build further.

Acknowledgements

This research was performed at the Sony Computer Science Laboratory in Paris and Tokyo. Additional funding came from the FP7 EU project ALEAR and the FP6 EU project ECAgents. The authors thank Martin Loetzsch for help in gathering the data. Martin Loetzsch and Simon Pauw contributed to technical issues related to the present paper. We thank them for their help. Also, we would like to express
our gratitude to Masahiro Fujita and Hideki Shimomura from Sony Corporation for supporting our experiments on the non-commercial Sony humanoid robots in Tokyo. We thank John Bateman for his extensive comments on earlier versions of this paper.

References


Appendix

The experiment reported in this chapter uses Fluid Construction Grammar as core formalism for grammatical processing (Steels, 2012). The details of this formalism are not important to understand the experiment, but we nevertheless provide an example of a lexical and grammatical construction for illustrative purposes.

The lexical construction is shown in Figure 12. Such a construction is the typical outcome of experiments in language learning as reported in Spranger (2012). The same construction representation is used unchanged for parsing and production. In parsing the construction is used to go from form to meaning and in production from meaning to form. The left side contains information on the semantic pole and the right side information on the syntactic pole. The top box provides the conditions in which this construction triggers and the bottom box the information that the construction adds if it has become active.

We see here that the meaning covered by this lexical item is a spatial category internally labeled ‘wekato’, and hence the meaning is equal to a bind-operation:

\[(\text{bind \ angular\text{-}spatial\text{-}category \ ?category-58 \ wekato})\]

The word associated with this meaning is the string “wekato”. The bottom box of the construction provides information about the semantic categories associated with this category, namely that it has the potential to be used as an angular-spatial-category and hence as a spatial category or more generally a category. The syntactic categories are entirely the same, meaning that in this phase they are purely semantic in nature - but may gradually evolve into purely syntactic categories such as adjectival.

Figure 13 shows an example of the kind of grammatical constructions that emerge in this experiment. This is a phrasal construction that brings three constituents together. The top-box left side contains again the conditional semantic pole and the right side the conditional syntactic pole. The bottom-box left side contains what the construction contributes to the semantic pole and the bottom-box right side what it contributes to the syntactic pole.

There are three constituents involved in building the phrase: One (labeled ?class-unit-25561) introduces an object class such as block, a second one (labeled ?sel-unit-17943) is a selector such as unique, and a third one (labeled
Figure 12. Example of a lexical construction for expressing a category similar to north and front.

?cat-unit-22249) is a spatial category. The phrase covers a complex network of semantic functions. The functions include identify-location, geometric transform, apply-selector, apply-class, and get-context. Some of the variables in this network will get bound through the lexical items that constitute the three constituents. On the syntactic side, we see that the construction introduces in production (or requires in parsing) an extra marker, in this case the string “bo”, in addition to the three constituents. The bottom-box contains the semantic constraints that the final target of the mental program has, in particular that it is a location, which is a point appearing as a sensory-entity.
Figure 13. Example of grammatical construction for marking the intrinsic conceptualization strategy (see Figure 8). The construction has three lexical constituents: a spatial relation, a selector and an object class. In production, the intrinsic strategy is expressed using the three lexical constituents plus the marker “bo” that is introduced by this construction. In parsing the construction fully recovers the intrinsic conceptualization strategies upon observing three compatible lexical items and the marker “bo”.