A Basic Emergent Grammar for Space

Michael Spranger

[draft - to appear in: Steels, L. (ed.) (2011) *Experiments in Language Evolution* John Benjamins, Amsterdam.]

Abstract

Grammar helps to avoid ambiguity and errors in interpretation. This chapter studies this claim with the domain of spatial language. It shows that agents equipped with language strategies which try to avoid ambiguity by expressing explicitly which conceptualization strategy is to be used to what context, are able to develop a grammatical communication system that optimizes their communicative success by diminishing cognitive effort and increasing expressive adequacy.

1. Introduction

An earlier chapter proposed and studied lexical language strategies that allowed the co-evolution of spatial terms and spatial categories in a distributed population of agents. Purely lexical communication systems are limited because while they enable agents to express spatial categories, they do not allow to encode how these categories are used to achieve a particular semantic function and they do not process the spatial context in which the categories need to be applied. Consider for example the following three utterances:

(1) *der* *linke* *Block*
the.NOM left.ADJ.NOM block.NOM

'The left block',

(2) *links* *des* *Blockes*
left.PREP.GEN the.DET.GEN block.GEN

'to the left of the block',

and

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

'to the left of the block',

They all contain the same lexical material and the same spatial categories but the way these categories are intended to be used is quite different. In Example 1 the spatial relation is used to further restrict the set of objects denoted by the noun, whereas in Example 2 the spatial category is applied to a landmark denoted by the noun. In Example 3 a determined noun phrase is followed by the category expressed as an adverb, which denotes a region that is used to modify the set of blocks. In contrast to Example 1, however, the spatial category is related to a covert landmark not a group-based relative reference system. If an agent in a *spatial language game* (**?**) is confronted with these utterances, he is supposed to process the context differently in each case.

Clues for the different processing of these utterances are in the grammatical structure of each of them. In the first case the word order and the morphology make it clear that the spatial category is used as an adjective modifier within a determined adjective noun phrase. In the second case, the category is expressed as a preposition which encodes that the determined noun phrase tail of the phrase is denoting the landmark to which the spatial relation is applied. In Example 3, the category is used as an adverb which can be seen du to the lack of a subsequent noun phrase. The use of the spatial category as adverb entails inferences about the type of the region, e.g., the region can be internal or external and there must be a covert landmark. Consequently, these three examples show how the German grammatical system provides valuable indications as to how to interpret each of these utterances.

This chapter investigates whether a population of agents equipped with a base lexicon could self-organize a set of grammatical markers that have the same function as the constituent structure, word order and morphology of German spatial language. The experiments are based on a reconstruction experiment using Fluid Construction Grammar, in which the relevant German spatial expressions were analyzed and a lexicon and grammar designed by hand. The relevant grammar has been discussed at greater length in (**?**). Using this reconstruction we can examine the precise role of grammar (section 2). Next we introduce the components of a language strategy (the invention, adoption, and alignment operations) that will allow the population to construct a spatial grammar, given a lexical system (section 3).

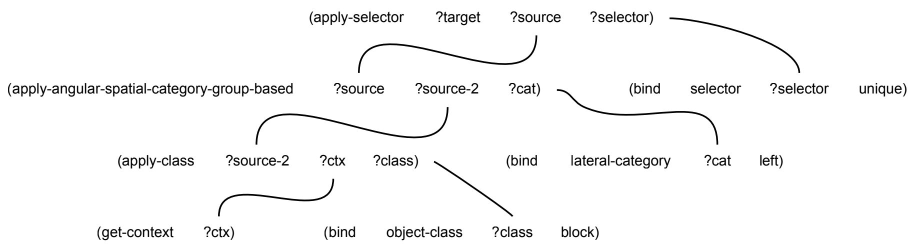

Figure 1. *Semantic structure underlying Examples 1 and 4. The semantic structure has three semantic entities which are linked in such a way that the set of blocks filtered by the primitive* `apply-class` *is further filtered by the spatial category* `left` *which is applied here to a group-based relative landmark . Finally, the selector* `unique` *is applied retrieving the best scoring element.*

2. The Role of Grammar

To examine the selectionist criteria that might push a population towards the introduction of a spatial grammar, it is helpful to consider what is left if one were to "remove" the grammatical cues from Examples 1 to 3. In principle, one ends up with a phrase like the following.

(4) *link block der*
left block the

This phrase is non-grammatical because it lacks proper German morphology and word order among other things. This phrase can be interpreted in many different ways. Since we represent semantics using IRL we will immediately make this point using IRL-networks to formalize what is meant by different semantics.

Three plausible spatial semantic interpretations of Example 4 are shown in Figures 1, 2 and 3. The semantic structures shown in these Figures formalize the intuitive notions as to what the difference is in the semantics of Examples 1 to 3 which are all utterances built using the same lexical material as Example 4. Most importantly, however, all three semantic structures are valid interpretations of Example 4 because Example 4 only constraints the space of possible interpretations by providing a set of semantic entities that need to be part of the semantic structure. These entities are part of all three semantic structures.

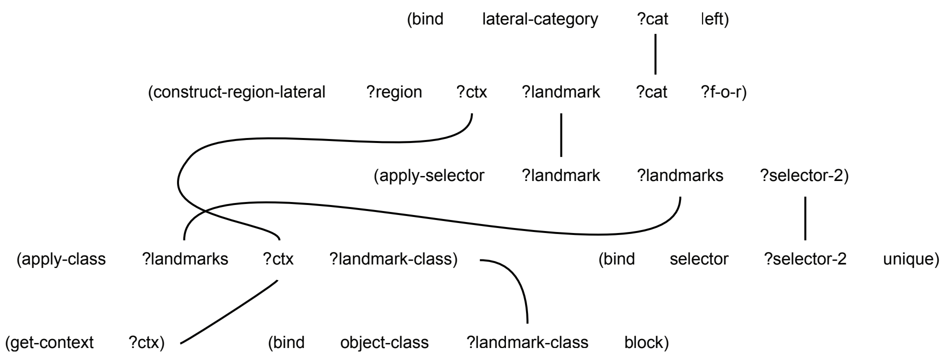

Figure 2. *Semantic structure underlying Examples 2 and 4. This semantic structure encodes how to construct a lateral region based on the spatial relation* `left` *using the landmark that is provided by the subpart of the structure singling out an object from the context by applying the object class* `block` *and the selector* `unique`.

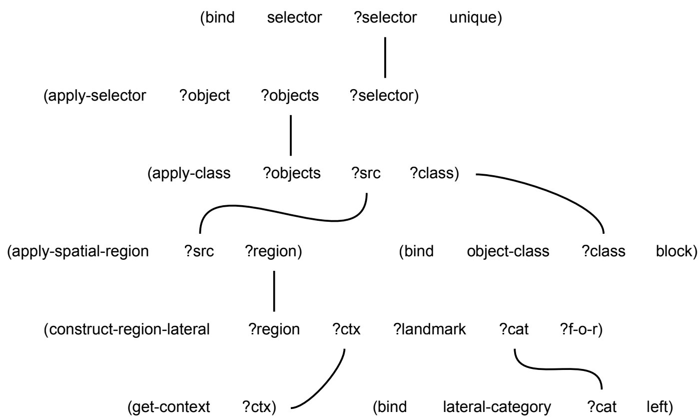

Figure 3. *Semantic structure underlying Example 3 and 4. The spatial relation* `left` *is used here to construct a region based on an unspecified landmark. The region is used to filter the context, followed by the application of the object class* `block` *to further filter the objects, followed by the application of the selector* `unique`.

To understand what information grammar provides in the formal framework of IRL we can examine the differences in interpretations. There are two important differences between these structures the first is related to the cognitive operations involved. For instance, semantic structure 1 involves the operation `apply-spatial-category-group-based` which is not found in 2 or 3.

The second type of difference can be seen when comparing structure 2 and 3 which differ primarily in how the cognitive operations are linked. In Figure 2 the landmark of the spatial region is given by the subnetwork consisting of `apply-selector` and `apply-class`, whereas in 3 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 3 the spatial region modifies the set of blocks. In 2 the block is the landmark of the region. In summary, grammar expresses which cognitive operations are part of the semantic structure of an utterance *and* how the operations are internally linked.

Removing the grammatical cues from Examples 1 to 3 increases the number of possible interpretations of these phrases. Consequently, grammar is related to *semantic ambiguity* which we define as different possible semantic structures underlying the same utterance (**?**). The increase in possible interpretations of a phrase impacts 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 topic.

- 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 4 shows the impact of grammar on communicative success in different spatial settings[1]. We compare two populations of agents. One is equipped with the full German locative grammar. The second population only is given lexical constructions. When an agent equipped with a purely lexical system speaks, he only conveys the semantic entities, e.g. spatial relations, object classes, determiners and discourse roles without explicitly marking their 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

1. We use the spatial setups and language games described in **?**.

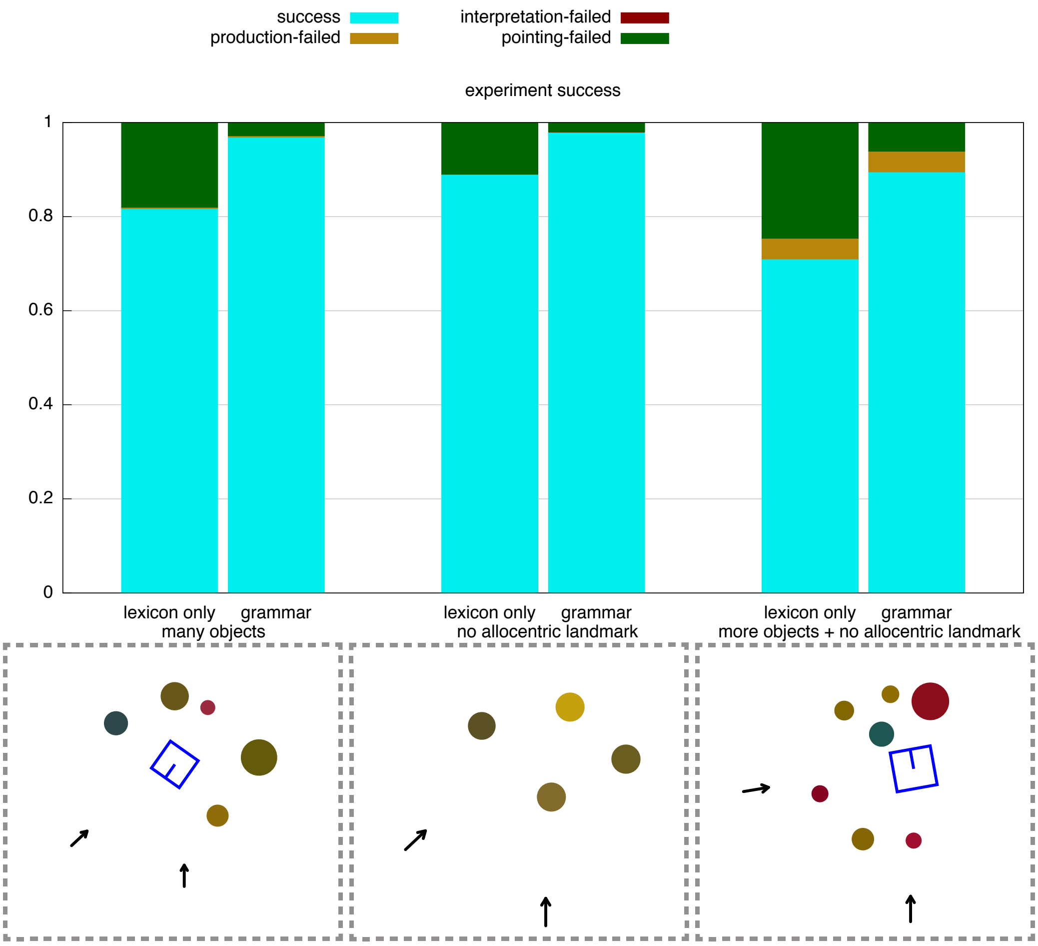

Figure 4. *The figure compares the performance of agents equipped with a purely lexical system with a population in which all agents operate a reconstructed German space grammar.*

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. Quite strikingly though, 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 agents are endowed with, which allows them 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 processing efficiency. Figure 5 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. For grammar the number drops to around 1.0, which means that in many cases an agents 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 claim 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, if perceptual deviation is minimal and if there are only few objects in the scene, the effect of grammar is less strong than in cases where viewpoints are different, perceptual deviation is strong and the number of objects is high (a fact that is demonstrated in Figure 4).

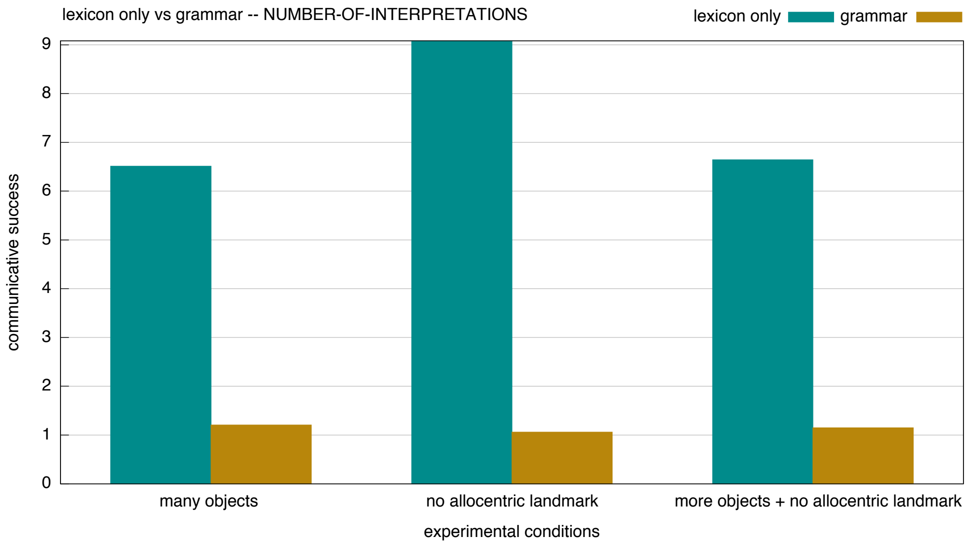

Figure 5. *This figure compares the number of semantic structures tried in interpretation for German locative phrases processed with and without grammar. Essentially this measures semantic ambiguity and shows efficiency in interpreting utterances. The figure shows the average number of interpretations (10000 interactions). In the grammar case the average is just barely above 1.0. We conclude that grammar greatly reduces the ambiguity in interpretation.*

But the relation between scene complexity, semantic ambiguity and communicative success is subtle. For instance, the results shown in Figures 5 and 6 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 6 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 of average interpretations of utterances (Figure 5) but the smallest drop in interpretation correctness. In other words, just because the number of interpretations is high does not necessarily mean that hearers will not correctly interpret. Consequently, communicative success drops only minimally in comparison to the other conditions.

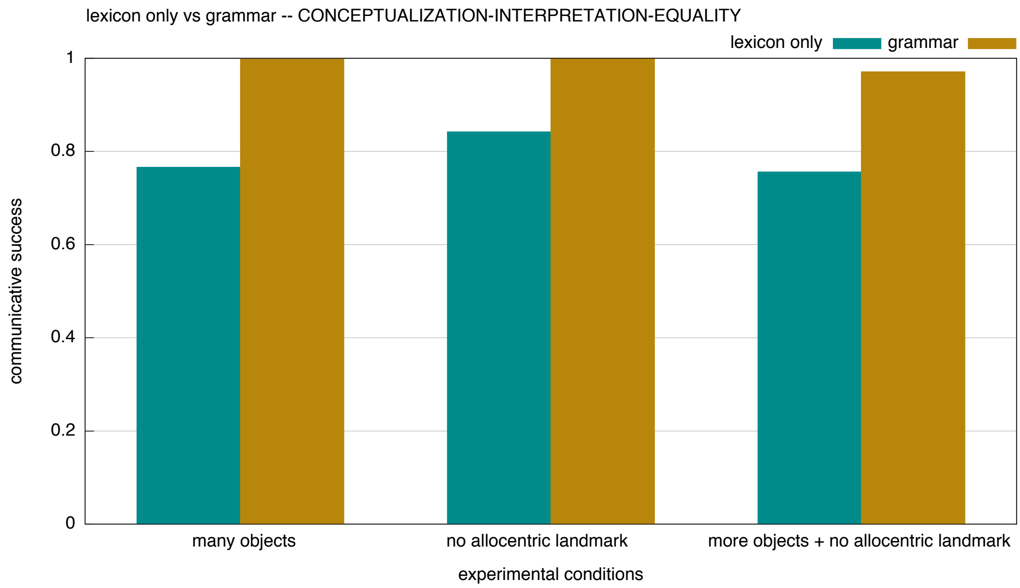

Figure 6. *This figure shows how equal the semantic structure recovered by the hearer is to the conceptualization strategy of the speaker The effect is compared for different sets of spatial scenes. If the semantic structures are equal the interaction counts as 1.0, if not as 0.0. The results show the 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 respect 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 impacts on the performance advantage grammar has over purely lexical systems. 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 than the ambiguity does not play a role in these conditions.

3. Emergence of Grammatical Markers for Space

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 allows agents to optimize their linguistic communication system by introducing grammatical markers to decrease cognitive effort and avoid interpretation errors. To focus on the main points being investigated here, we do not look at the full blown German locative system but only a limited set of conceptualization strategies (in particular the one shown in 7) using a grammar which is deliberately kept simple.

Concretely we scaffold agents with four conceptualization strategies: intrinsic (illustrated in 7), absolute, relative from the perspective of the speaker and relative from the perspective of the hearer. Agents are equipped with four angular spatial relations, all of which are modeled after “vor” (front), “hinter” (back), “links” (left) and “rechts” (right) with the difference that they behave like “über” (up) and partake in intrinsic, relative and absolute conceptualization strategies (see **?** and **?** for the difference in intrinsic, relative and absolute conceptualizations).

Agents operating only lexical constructions will express themselves always in three word utterances. The utterances consist of the spatial relation, as well as a determiner and a noun. All four strategies operate the same set spatial relations and all apply the frame of reference to an allocentric landmark which can be marked lexically. For any utterance the three words alone never distinguish between the four conceptualization strategies given to agents, but rather a hearer always has to try all four strategies, in order, to retrieve the most likely topic.

Another set of agents is equipped with grammatical constructions, which allow them to mark the conceptualization strategy. Figure 8 shows a grammatical con-

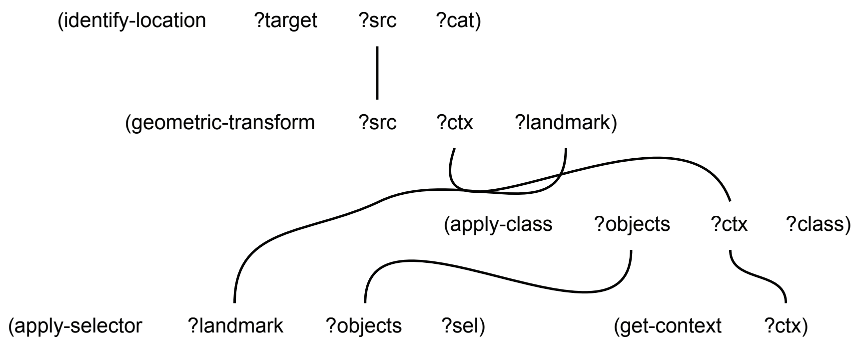

Figure 7. *IRL-network of the intrinsic conceptualization strategy. A corresponding construction for expressing this strategy is shown in Figure 8.*

struction that marks the intrinsic strategy (see Figure 7) with the marker "bo". Consequently, if an agent with grammar uses the intrinsic conceptualization strategy, he constructs an utterance such as "der bo linke block" (with word order being irrelevant) Subsequently, in interpretation the hearer of this utterance parses a complete IRL-network and he is not required to additionally process other conceptualization strategies to find the topic of the phrase.

Figure 9 compares the performance of lexical agents that only operate lexical constructions with agents that grammatically mark the conceptualization strategy they are using. We can observe similar positive effects of grammar, both on processing and communicative success, as already observed for the complete German locative system, showing that the simplified morphological design pattern used here adequately serves the same function.

3.1. Invention and Alignment Operators

The next question is by what kind of invention, adoption, and alignment operators the agents might be able to introduce and coordinate these grammatical markers. We follow the same approach as used in earlier chapters: Agents monitor their performance, diagnose problems and attempt to repair diagnosed problems for the future by changing their lexicon or grammar.

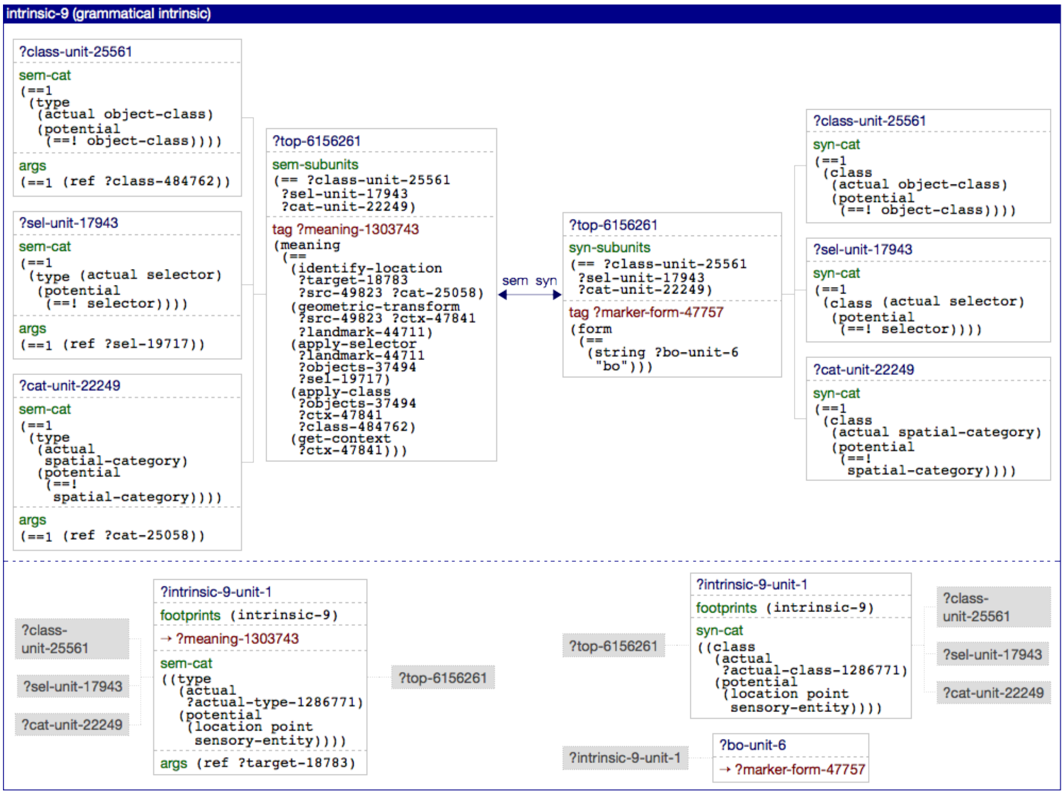

Figure 8. *Example of grammatical construction for marking the intrinsic conceptualization strategy (see Figure 7). The construction has three lexical constituents: a spatial relation, a selector and an object class. In production, the intrinsic strategy is expressed using the three lexical constituents 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".*

Let us first consider invention. Before speaking, the speaker *re-enters* his utterance. He parses the utterance himself, in order, to diagnose problems that might occur. If for instance the utterance is ambiguous, it might be interpreted as referring to different objects. If the speaker diagnoses this problem, a corresponding

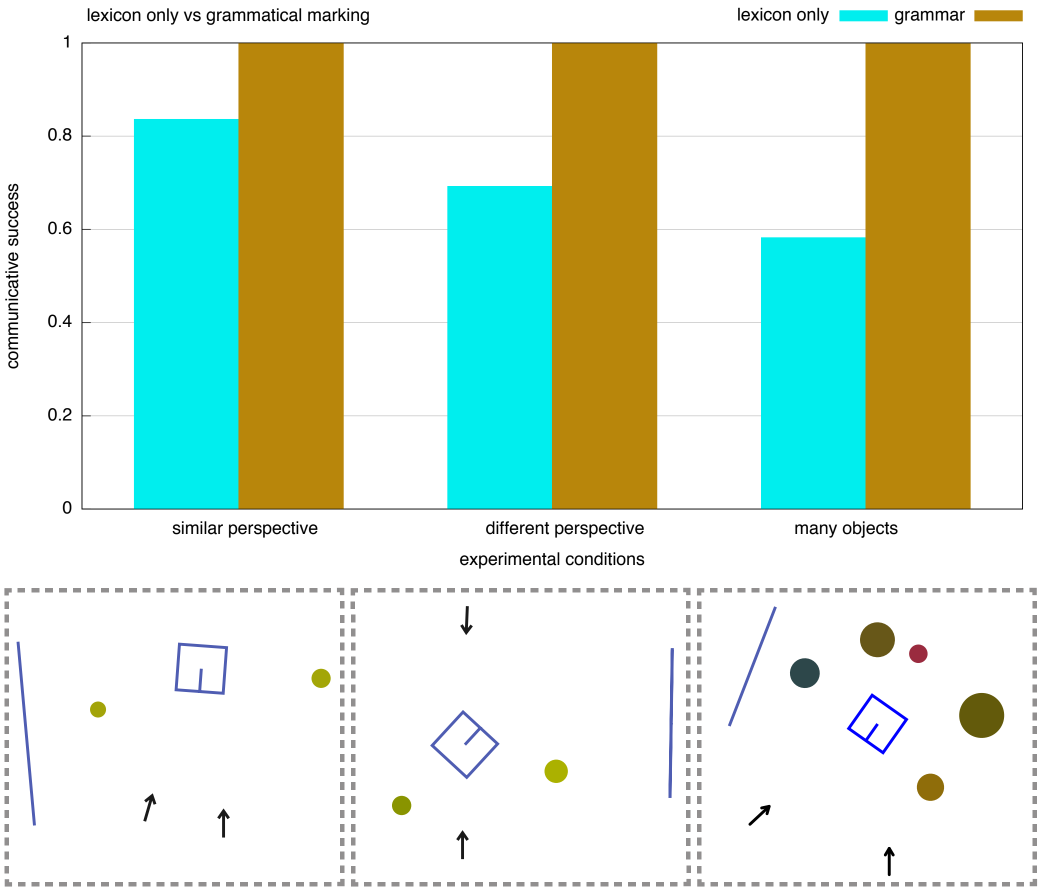

Figure 9. *Comparison of purely lexical agents equipped with four angular conceptualization strategies. Half of the strategies are dependent on perspective, hence, the impact of changing the perspective of robots on the scene is a major impact, another is the number of objects in each scene.*

repair strategy triggers which tries to solve the problem by inventing a grammatical marker. The following is an overview of the invention operator

1. *Speaker encounters multiple target objects in re-entrance.*

- Diagnostic: In re-entrance the speaker tests the interpretations of what he is about to say to the hearer. If there are multiple interpretations of the utterance leading to different possible interpreted topics, this is diagnosed as a problem.
- Repair: The speaker invents a grammatical constructions such as in Figure 8 which expresses the parts of the meaning that are not expressed by lexical items.

When a new marker was invented, the hearer will inevitably encounter the new marker. A second repair strategy takes care of learning unknown markers.

1. *Hearer encounters an unknown string s*
   - Diagnostic: When the hearer does not know a term.
   - Repair: The hearer waits until the topic of the conversation is successfully established (pointing of the speaker or correct pointing of the hearer). He then builds a construction that links those parts of the meaning which were not parsed with the unknown term *s*.

At the end of the interaction both interacting agents have a grammatical construction that links the new marker to the conceptualization strategy both have used in production and re-production respectively.

Next to invention and adoption, agents need alignment strategies. There are two goals for alignment strategies. First, the hearer might adopt the marker such that it links to a different conceptualization strategy then the one used by the speaker. In this case the alignment operators have to be such that one of the wrong mappings dies out over time. The second goal is that if there are multiple markers occurring in the population for the same conceptualization strategy, then the agents of the population should come to an agreement as to which marker to use for that particular strategy. Alignment for grammatical constructions is the same as for lexical constructions. Successfully used constructions are rewarded, unsuccessfully used constructions punished. Additionally, competitor constructions are punished which are those constructions that could have been used in production but were not used because another construction covering the same semantic structure was applied and has a higher score. This implements *lateral inhibition* (**?**) dynamics for grammatical constructions.

3.2. Experimental Setup and Results

We test the performance of the proposed invention, adoption and alignment operators on different sets of spatial scenes. Agents are equipped with four spatial relations, one determiner and three object classes together with lexical constructions for expressing these items. In total agents are given eight lexical constructions for the four spatial relations, the determiner and the three object classes (robot, box and block). Moreover, agents are given the four conceptualization strategies discussed earlier. The task for agent is then to develop a system that allows them to increase their success from the lexicon only baseline condition to 100% success.

Figures **??** and **??** shows the dynamics of development for populations operating the invention and alignment operators as well as the lexical constructions and the semantic entities discussed. Agents are able to develop successful grammatical marking systems given the need to disambiguate semantic structure in all three environmental conditions considered. In all cases agents develop a grammatical marker system consisting of four markers marking the four conceptualization strategies. We can also see that essentially in all conditions markers are necessary for disambiguation. It is just the number of contexts that require disambiguation that drives development of the marker system. Clearly this number is low in the case of the *similar perspective* condition,

4. Conclusions

This chapter showed how a set of grammatical markers that signal which semantic function spatial terms are intended to perform as well as their context of application can self-organize in a population of agents endowed with a lexical spatial system. The main point is that semantic ambiguity (multiple-target-entities) occurs when semantic function is not expressed and this generates a search space that the agent needs to explore by examing the world model and it leads to the risk of interpretation error when the wrong interpretation is chosen.

Of course the notion of grammar used in this section is weak in many respects. It is at best an abstraction that tries to preserve some structural properties, such as the relation of grammar to structural properties of semantic structure, as well as the relation to cognitive operations. Nevertheless, this section shows why grammar is important on a fundamental level and how this purpose of grammar can be used for self-organization of such a system. The argument was backed up by experiments in which the grammar part is removed, reducing agents to use merely lexical systems and how it can grow back given the right set of invention and alignment operators. In

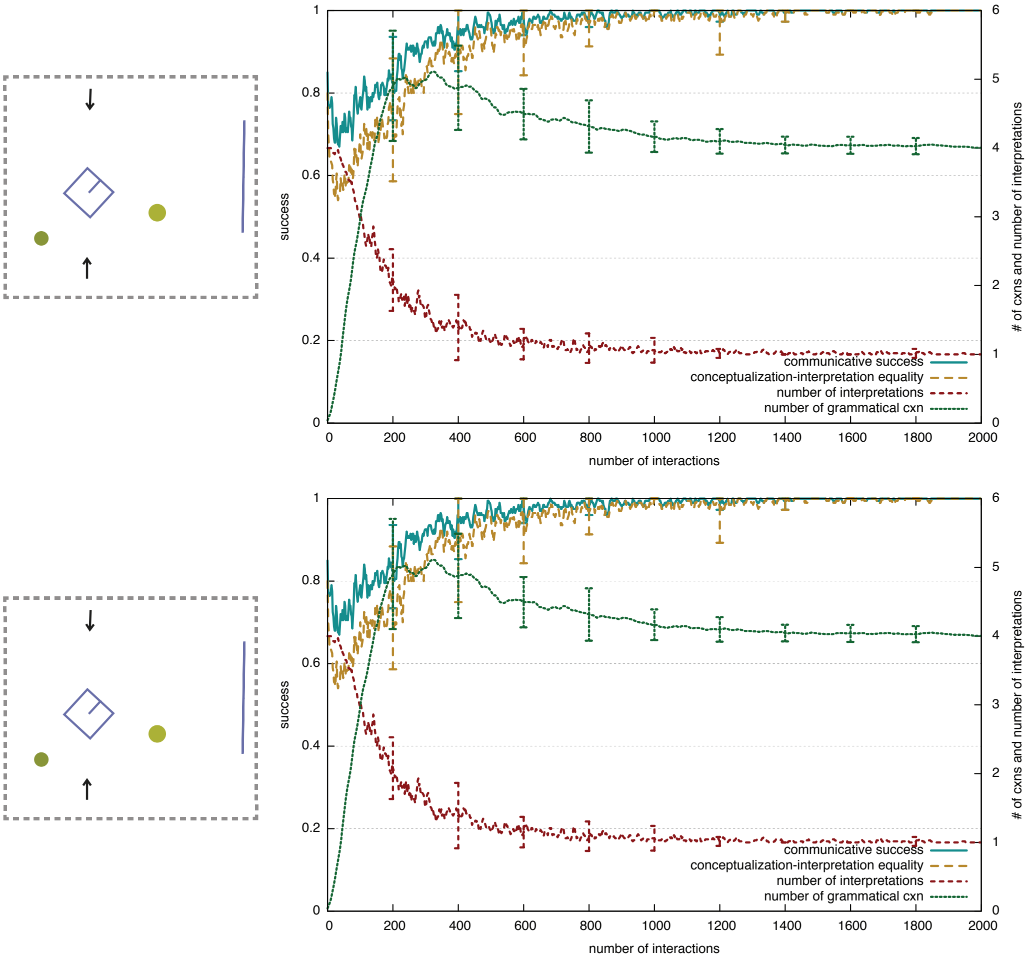

Figure 10. *Development of grammatical markers over time. As the number of grammatical markers in the population increases the number of interpretations per phrase drops. In all three conditions agents self-organize a grammatical communication system and are able to increase success from the base line of purely lexical success to more or less 100%. At the same time as the system of grammatical markers develops, we can measure a drop in tested interpretations (diminishing cognitive effort) and an increase in correct interpretations (increasing communicative success).*

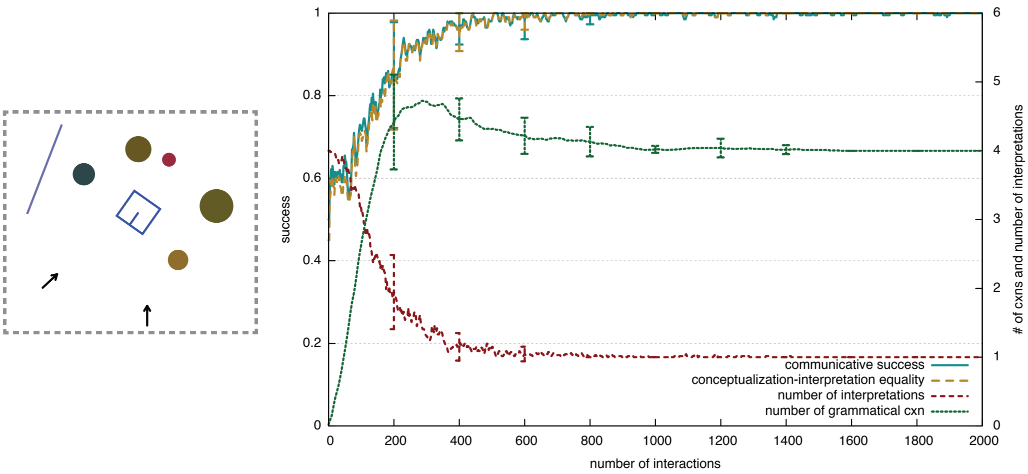

Figure 11. *Development of grammatical markers over time in the third experimental condition.*

principle, this kind of argument, therefore, shows less how a particular grammatical system can evolve because much of this intelligence was provided in the learning and alignment operators, but it shows the necessary conditions for the emergence of grammar.

Acknowledgements

This research was performed at the Sony Computer Science Laboratory in Paris. The author thanks Martin Loetzsch, Luc Steels and Remi van Trijp for help in data-gathering and Luc Steels in help for writing the present paper. We thank Hideki Kimomura for supporting experiments on the non-commercial Sony humanoid robots in Tokyo. The research was partially funded by the EU FP7 project ALEAR.