Co-Acquisition of Syntax and Semantics — An Investigation in Spatial Language

Michael Spranger[1] **and Luc Steels**[2]
1) Sony Computer Science Laboratories Inc., Tokyo, Japan, michael.spranger@gmail.com
2) ICREA, Institut de Biologia Evolutiva, Barcelona, Spain, steels@arti.vub.ac.be

Abstract

This paper reports recent progress on modeling the grounded co-acquisition of syntax and semantics of locative spatial language in developmental robots. We show how a learner robot can learn to produce and interpret spatial utterances in guided-learning interactions with a tutor robot (equipped with a system for producing English spatial phrases). The tutor guides the learning process by simplifying the challenges and complexity of utterances, gives feedback, and gradually increases the complexity of the language to be learnt. Our experiments show promising results towards long-term, incremental acquisition of natural language in a process of co-development of syntax and semantics.

1 Introduction

Here is an example of a locative spatial phrase from English.

(1) The block left of the box from your perspective.

A sentence such as this can be used by a speaker to single out an object in the environment and draw attention of the listener to that object. To reach this goal the phrase includes a number of locative spatial components. There is a *locative spatial relation* "left", *a landmark* "the box" and *a perspective* "from your perspective". The components are put together to signal to the listener exactly how to identify the referent of the phrase.

Spatial language is of huge importance for communication. Despite being so cognitively central and perceptually determined, spatial language turns out to be largely culture-specific [Levinson, 2003]. This not only true for the lexicon and grammatical aspects [Svorou, 1994] but also for the conceptual structures a language affords. Tzeltal, a Mayan language impressively demonstrates this through its sole reliance on absolute spatial relations such as uphill/downhill and its lack of projective categories such as front/left [Brown and Levinson, 1993]. To autonomously learn a spatial language, therefore, requires not only to learn the expression of spatial relations but also to acquire the underlying conceptual repertoire.

This paper is part of a larger effort on understanding the developmental origins of language by emulating stages in development similar to children using cognitive linguistics and construction grammar, specifically Fluid Construction Grammar as foundations. Its developmental perspective makes it part of Developmental A.I., which has made huge steps forward the past decade through detailed models of sensorimotor skill learning [Asada *et al.*, 2009]. However, only a few models have dealt with the acquisition of spatial language. There is some work on the learning of grounded spatial relations [Spranger, 2013; Bailey *et al.*, 1997] in tutor-learner scenarios, but that has only focussed on spatial categories and not on grammar. There have also been attempts on development of grammar, e.g. [Saunders *et al.*, 2009], but those approaches do not go beyond very early stages of grammar development, and typically neglect the semantic aspects of grammar. A notable exception is [Alishahi and Stevenson, 2008] which is a non-grounded model of learning argument structure constructions.

In the robotics community, spatial language plays an important role for commanding robots, describing the environment and referring to objects in, for example, industrial task settings. Often then hand-crafted systems [Winograd, 1971] or supervised machine learning techniques [Chen and Mooney, 2010] are used. More often than not the semantics of the language is fixed. The language component maps utterances to the given symbolic meaning space, which allows statistical techniques from machine learning to be applied [Matuszek *et al.*, 2010]. A notable exception is [Tellex *et al.*, 2013], which learns a probabilistic graphical model from unaligned text, scene pairs and which can also learn to ground the meaning of words.

Lastly, in the Cognitive Science community computational models of spatial lexicon acquisition [Regier, 1996] have been very influential. Importantly, while often not tested with real robots, these models give quantifiable estimates of the impact of certain strategies such as using cross-situational statistics [Frank *et al.*, 2008; Siskind, 1996] and biases [Saunders *et al.*, 2011; Griffiths *et al.*, 2010].

Our work is based on many of these previous insights but is different in its focus on grounding and the incremental co-acquisition of spatial grammar and complex semantics.

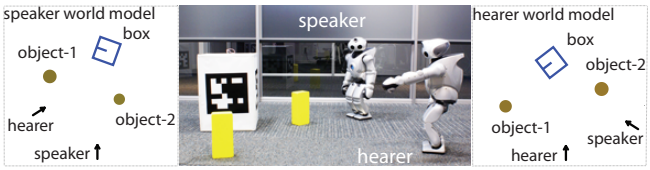

Figure 1: Spatial language game set-up. The images left and right show the situation model as perceived by each robot.

2 Embodied Language Games

We use *language games* to study the acquisition of spatial language [Steels, 2001]. Two robots (one is the tutor, the other the learner) are interacting in a shared environment and are trying to draw each others attention to objects in the vicinity, using language. The set-up is shown in Figure 1. The environment consists of a number of objects (represented in the graphical representation of the situation model as circles), boxes (rectangles) and interlocutors (arrows). The vision system of each robot tracks objects in the vicinity and establishes a model of the environment with real-valued distances and orientations of objects with respect to the body of the robot. The environment is open-ended. Objects and boxes are added or removed and their spatial configuration changed. Moreover, robots are free to move around. For the purpose of evaluation we recorded more than 1000 spatial scenes with different numbers of objects (up to 15 objects) and different configurations of objects.

1. Each agent perceives the scene
2. The speaker selects an object (further called the topic T) from the situation model. The speaker tries to find a meaning (this can involve spatial relations, landmarks and perspective) which discriminates T from other objects in the scene. Subsequently, the speaker produces an utterance for this meaning.
3. The listener parses the utterance and tries to interpret the (potentially partial) meaning of the observed utterance to identify possible topics in the scene. The listener then points to the topic he thinks is the most likely interpretation of the phrase.
4. The speaker checks whether the listener points to T. If the listener pointed correctly, the game is a success and the speaker signals this outcome to the listener.
5. If the game is a failure, then, depending on the tutoring strategy, additional things may happen. The tutor can point to the object he had in mind, the topic if he is the speaker, or the thing he understood to be the topic when he is the listener. Alternatively, a learner may point to the topic, and the tutor might say how he would refer to the topic.

3 Representing and expressing Spatial Meaning

We are using a particular formalism called *Incremental Recruitment Language* (IRL, [Spranger *et al.*, 2012]) to represent the *procedural semantics* [Haddock, 1989] of spatial

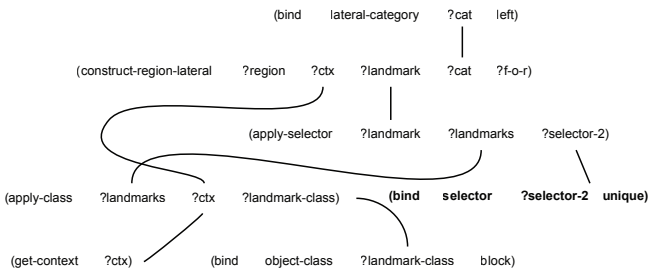

Figure 2: IRL-program Example 3.

utterances. Figure 2, for instance, shows the IRL-program (meaning) of the phrase "left of the block". The structure contains pointers to concepts and spatial relations in the form of *bind statements* (in bold), as well as a number of *cognitive operations*. For example, `construct-region-lateral` constructs a region representation. Cognitive operations and bind statements are linked using variables (which are symbols starting with ?). For instance, the variable `?lm` links a subpart of the IRL-program identifying "the box' to the landmark input slot of the operation `construct-region-lateral` thereby capturing the fact that "the box" should act as the landmark to the region.

Spatial relations: English locative spatial relations can be broadly categorized into three different classes.
Proximal categories such as "near" and "far" rely on proximity to some particular landmark object.
Projective categories are categories such as "front", "back", "left" and "right". These categories are primarily angular categories signifying a direction. The direction can come from the reference object itself (intrinsic) or can be induced by the observer or some perspective (relative frame of reference [Retz-Schmidt, 1988]).
Absolute categories such as "north", "south", "east" and "west" which rely on a compass directions, with the pivot direction to the magnetic north pole. Other absolute systems rely on features of the environment to determine the layout of the angles [Brown, 2008]. In the experiments discussed here, the wall marker is used as a global direction on the scene.

We represent spatial categories using a similarity function [Herskovits, 1986] based either on a prototypical angle (for absolute, projective) or distance (for proximal) enveloped by an exponential decay:

$$\mathrm{sim}_a(o, c) \quad := \quad e^{-\frac{1}{2\sigma_c}|a_o - a_c|}$$

where o is some object, c the category, a_o the angle to a particular object o and a_c is prototypical angle of category c. Importantly, angular and proximal distances are always defined relative to a coordinate system origin. By default this is the robot observing the world.

Cognitive operations: Agents can use spatial relations (and other concepts) in IRL-programs combined with different cognitive operations:
Set operations such as picking the highest scored member of a set etc., which are important for dealing with determiners such as "the".
Categorization operations take a set as input and score objects

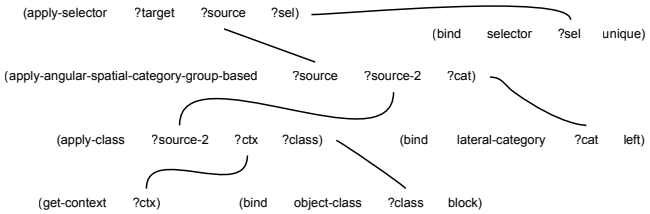

Figure 3: IRL-program for Example 2.

according to some similarity functions defined by categories. Examples are `apply-class` and `apply-category`.
Mental rotations: are implemented as linear algebra operations that transform a feature space such as angle and direction to another point of origin, e.g. `geometric-transform`. These operations also handle different frames of reference (intrinsic, relative and absolute).

An important insight from cognitive linguistics is that there is a deep connection between semantics and syntax. The following gives two examples from English to highlight this.

(2) The left block.

(3) left of the block .

Both phrases consist of the same lexical material (the, block and left) but their grammatical structure and their meaning structure is quite different. In Example 2 the spatial relation is used as modifier on the set of objects denoted by the noun, whereas in Example 3 the spatial category is applied to a landmark denoted by the determined noun phrase. In Example 2 the spatial relation refers to a group-based relative reference system [Tenbrink and Moratz, 2003]. Importantly, these differences are signaled by word classes. When "left" is used as a preposition, then it is used to construct a region. When "left" is used as adjective, then the group-based reference operation is needed. The word order of the utterance and grammatical markers such as "of" communicate how these different cognitive operations are linked.

4 Acquisition of Spatial Categories

The first step in spatial language learning is the acquisition of the spatial relations. Acquisition of a category is a two-step process. It starts with the learner encountering an unknown word in a particular communicative situation. To *adopt* the word, the learner stores it together with an initial estimated category as meaning. The information available to the learner in a single interaction is typically insufficient for a good estimate of the spatial category. The learner will therefore have to integrate information from subsequent interactions in which the new word is used to *align* better to the tutor.

Categories are initially encountered in a particular interaction using the following operation:
Listener encounters unknown spatial term s
Problem: Listener does not know the term (step 4 fails).
Repair: Listener signals failure and the speaker points to the topic T. Subsequently, the listener constructs a spatial category c based on the relevant strategy (projective, proximal or

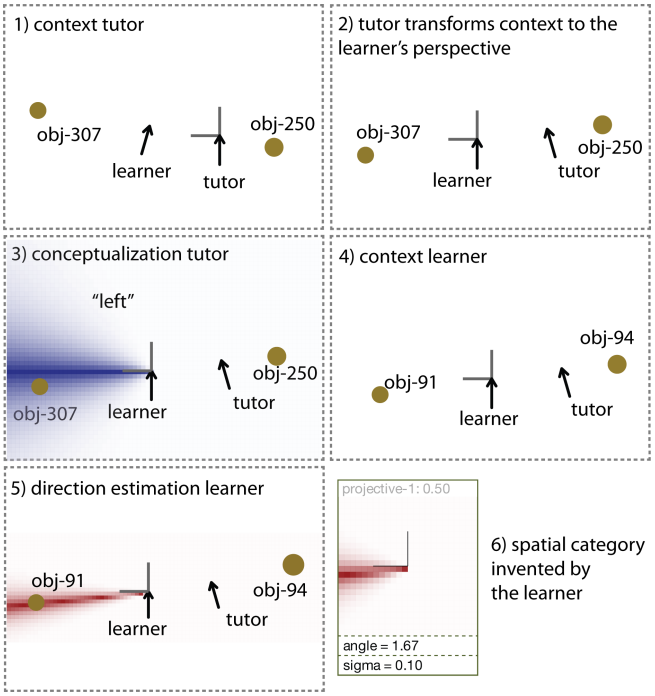

Figure 4: Adoption of an unknown category label by a learner agent in interaction with a tutor agent. The tutor starts by conceptualising the topic object in his world model (image 1). Here, `obj-307` (`obj-91` learner's world model) is chosen as topic. The tutor conceptualises a meaning for the topic from the perspective of the learner (image 2). The tutor finds the category `left` associated with the word "left" to be most discriminating (image 3). The speaker utters the word to the learner (context learner image 4). The learner does not know the word and the interaction fails. After the speaker pointed to the topic, the listener can adopt the string and connect it to the newly invented projective category `projective-1`.

absolute) and the topic pointed at (see Figure 4). Additionally, the listener invents a mapping associating c with s.

New words are always adopted in a particular interaction. Angle and distance prototypes are based on the particular distance and angle of the topic of the interaction to the learner. These are never exactly the same distance and angle of the category used by the tutor. To align his representation of the category of the tutor over time, the learner incrementally updates the category. For this he keeps a memory of past distances and angles. After each interaction the learner updates the prototype by averaging the angles (or distance) of objects in the sample set S of experiences of the category. The new prototypical angle a_c of the category is computed using the following formula where a_o is the angle of sample o.

$$a_c = \operatorname{atan2}\left(\frac{1}{|S|}\sum_{o\in S}\sin a_o, \frac{1}{|S|}\sum_{o\in S}\cos a_o\right)$$

$$\sigma_c' = \sigma_c + \alpha_\sigma \cdot \left(\sigma_c - \sqrt{\frac{1}{|S|-1}\sum_{o\in S}(a_c - a_o)^2}\right)$$

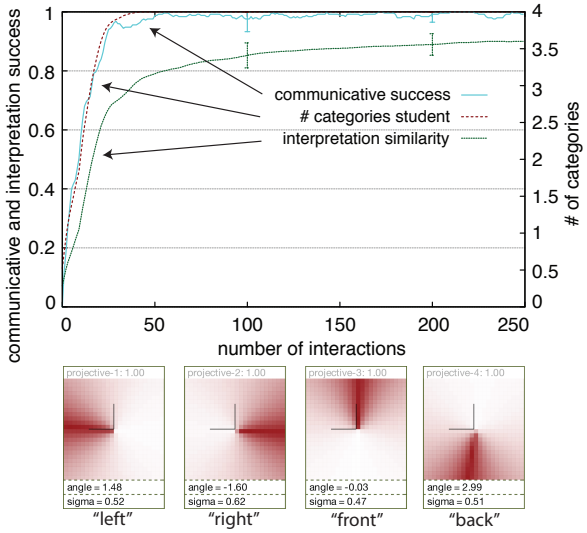

Figure 5: Dynamics of lexicon acquisition.

The new σ value σ' which describes the shape of the applicability function of the category is adapted using the following formula. This formula describes how much the new σ_c of the category c is pushed in the direction of the standard deviation of the sample set by a factor of $\alpha_\sigma \in]0, \infty[$.

We have tested the learning operators in experiments with a population consisting of one English category tutor and one learner. Figure 5 displays aggregated dynamics of 25 experimental runs showing the acquisition of projective categories (similar results exist for proximal and absolute categories). The learner quickly reaches communicative success. After roughly 25 interactions, all categories and their corresponding strings have been adopted. In the remaining interactions the alignment operator drives the *interpretation similarity* towards 1.0 (which is the highest value and signifies total overlap between the categories of the tutor and the learner). The bottom figure shows the categories acquired by the learner in one particular acquisition experiment.

5 Learning Spatial Grammar

In this section we focus on the strategies for learning spatial syntax. We assume, for now, that both agents share the IRL-programs and categories for conceptualizing and interpretation. This is obviously a strong assumption but it makes it possible to describe the learning of constructions in isolation. Later on we take away this scaffold. The learner starts out with no syntactic knowledge (no words or knowledge about phrase structure).

Listener encounters unknown spatial phrase s
Problem: Listener does not know the phrase or some part of the phrase (step 4 fails).
Repair: Listener points (if he can still interpret the phrase given the context t) or signals failure and the speaker points to the topic t. In any case, the listener will have constructed an IRL-program for the topic. The learner invents a mapping from the complete IRL-program to the complete phrase (or parts thereof).

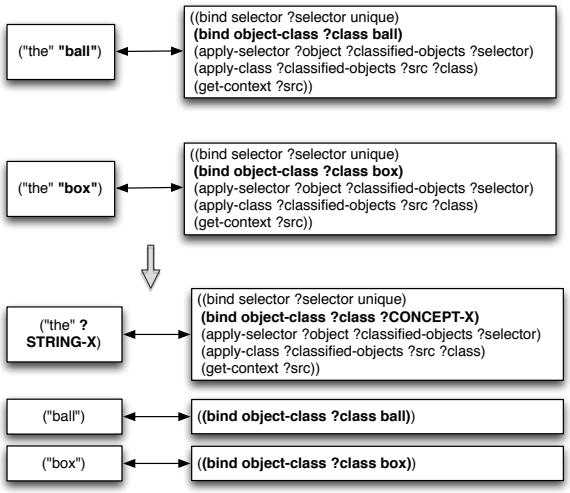

Figure 6: Schematic of an item-based construction (and two more lexical constructions) invented by the learner through reasoning over holophrases that have been heard before. The input constructions differ in semantics and in the form in a single item, which allows the learner to make a structural inference and split up the existing constructions.

Once the learner has acquired enough exemplars, he tries to extract more abstract constructions. Suppose the learner first hears the determined noun phrase "the box". Initially this will allow him to successfully produce and interpret that exact phrase. Upon hearing another example of a determiner and a noun "the ball", the learner can now deduce that likely he can build phrases of the form "the X" where X is something else namely a particular concept. The learner then invents an item-based construction and other constructions by breaking up the holophrases (see Figure 6 for a graphical explanation).

Once an item-based construction and its associated more lexical constructions have emerged, they are in competition with the holophrase constructions since they cover the same communicative situations (same meanings and phrases). The learner in production and interpretation knows this and so a competition takes place in the learner between the new constructions and the old holophrases. Setting up the right alignment dynamics for the learner can eliminate the holophrase constructions. Initially the learner will choose the new constructions over the older ones. Keeping track of how successful they are and which constructions compete. Punishing competing constructions after every interaction leads to a forgetting of the holophrase constructions over time. Once multiple item-based constructions are learned, they can be further broken up using the exact same learning operator. More and more abstract constructions will emerge with more possible arguments until finally something similar to phrase structure constructions emerges.

We can test the learning by running multiple tutor-learner simulations. The tutor can express a number of meanings and phrases. In total we see that 546 phrases are used by the tutor (on this particular data set) including determined noun phrases, e.g. "the block" or "the box", more complex adjec-

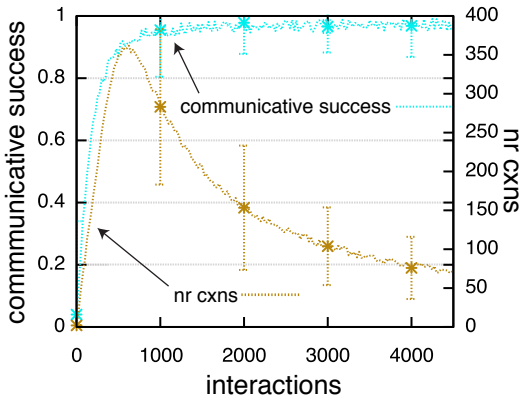

Figure 7: Dynamics of grammar learning given the learner knows all possible IRL-programs and categories.

tive noun phrases, e.g. "the left block", and very complex phrases such as "the block left of the box from your perspective".

Figure 7 shows the overall dynamics of learning spatial grammar (for 100 experiments). In all 100 experiments, the learner learns to be successful in communication after roughly 350 interactions (average 80% success). At this point the learner has been exposed to an average of 120 utterances (for about 550 total possible utterances used by the tutor), which is before the learner has seen all possible utterances, but enough to be successful for that stage. Initially the tutor chooses to expose the learner to simple scenes. This makes the learner immediately successful. Over time the tutor increases the complexity of the environment and the language needed to cope with the environment. That's why the learner does not reach 100% success but keeps learning.

At the same time, we can see an overshoot of the number of constructions in the inventory of the learner. This is due to an initial rise of holophrases, which later slowly die out, being overtaken by more item-based constructions and the emergence of lexical constructions (bind statements only meaning), The lexical constructions give already a lot of information and often allow the learner to successfully interpret the phrase given simple environments. Later functional constructions (handling a single cognitive operation or a set of them) and more phrase structure like constructions (only mapping variable linkings to word order) emerge. Slowly memory is consolidated. Typically, the learner does not end up with 46 constructions (which is what the tutor is using). A few more survive because the learner has not seen some enough examples of complex constructions. Importantly, the learnt system is fully productive more or less from the start and the learner can parse and make utterances he has never seen before.

6 Co-acquisition of Semantics and Grammar

The last piece of the puzzle for learning spatial language is the development of complex semantics itself. We represent complex semantics using IRL-programs such as the one in Figure 3. Consequently, learners need a way to automatically assemble new IRL-programs. IRL comes with such mechanisms. Starting from a pool of cognitive operations, IRL can put together IRL-programs in a process of automatic programming, where new programs are created and tried out by the learner. So for instance, initially the learner might use a simple `apply-category` operation similar to the one used in category learning. Later the `apply-category` operation can be combined with more complex operations such as mental rotation `geometric-transform`. See Figure 8 for a schematic explaining the process.

The process of creating IRL-programs is a heuristics guided search process based on communicative intentions shared by the interlocutors. For instance, the agents try to be most discriminative in their choice of feature channels, perspective, landmark etc. Also, the learner extends its repertoire of IRL-programs in a particular spatial context given a particular utterance of the tutor. For instance, the learner might be able to detect certain lexical items when he is trying to guess the meaning of an utterance. This partial information is used to constrain the space of possible IRL-programs.

Similarly to constructions and categories, the learner track the success of the IRL-programs he is using. Successful structures are retained for future interaction. Unsuccessful structures might be forgotten. Each IRL-program has a score and the learner increases the score of the IRL-program when it was used successfully and otherwise he decreases the score. If the score falls too low, the structure might be removed.

In order to learn more and more complex semantics starting out from simple ones, the communicative setting needs to be controlled so as to allow the learner to go through a slow process of incremental complexification. Similar to the previous experiments, the tutor takes on the role of simplifying the language up to a point, that the learner can establish first knowledge of concepts before acquiring more complex semantics. Roughly, the tutor will push the learner to go through the following stages.

Concept learning Initially, the tutor starts with simple environments that allow the learner to start with simple IRL-programs and conversely simple language (similar to the category learning described earlier).

Simple phrase learning In the second stage, the language becomes more complex and simple phrases such as the "the block" etc are used. Here first item-based constructions are learned and some simple phrases emerge.

More complex phrases In the third stage, complex phrases such as "the block left of the box" require the learner to start incorporating mental rotation operations.

More complex spatial semantics and phrases Here the tutor will start using spatial adjectives and relative frames of reference utterances. The learner has to come up with semantics for group-based reference and perspective.

Initial experimental results (combining all the mechanisms for concept, grammar and semantics learning) are very promising for our approach. See Figure 9 and 10 for an overview of the dynamics. The learner transitions incrementally through learning more complex utterances and semantics based on the tutor-guided acquisition process. This has

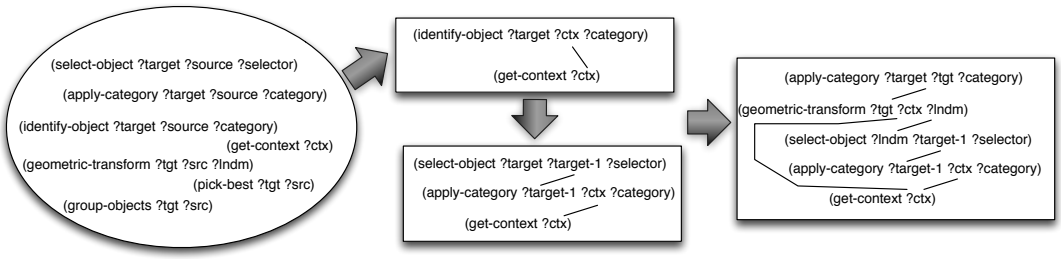

Figure 8: Recruitment of cognitive operations to form new meaning in the form of IRL-programs. Left, the pool of operations from which to build new structures. Middle top, a new program configured from the pool (this is what is used to learn the first categories). When spatial categories have been acquired new structure might be build based on an increase in complexity of syntax (here the part of the meaning of phrases like "the block"). Right, more complex structure emerging later in the acquisition (similar to "left of the box") process.

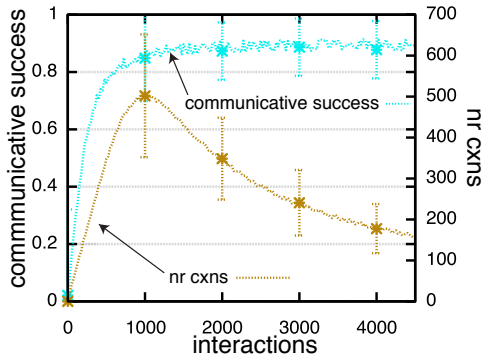

Figure 9: Dynamics of grammar learning when the learner simultaneously acquires IRL-programs and categories.

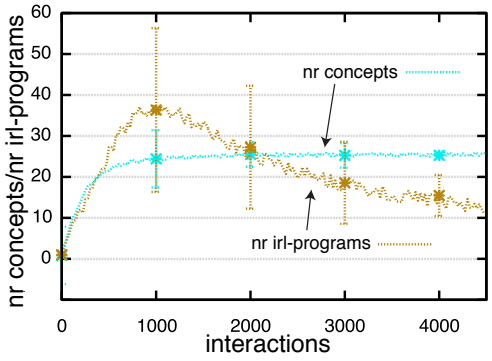

Figure 10: Development of the learners concepts and semantic structure for the same experiments as in Figure 9.

little impact on his success (because the environment is controlled by the tutor). However, the learner manages to acquire the 25 concepts needed and also manages to construct useful IRL-programs (the tutor uses 7). Interestingly, even while becoming more successful in more and more complex utterances, the learnt system stays quite messy. All sorts of meanings and constructions not used are still lingering around in the memory of the agent. One aspect of future work is thus to investigate adequate strategies of memory consolidation and forgetting.

7 Conclusion and Future Work

The experiments developed in this paper show that it is possible to setup scenarios involving tutors and learners that allow a fruitful study of the incremental development of language in artificial agents. We have used these setups to quantify the impact of tutoring strategies, the interaction scripts, the complexity of the environment, and the power of learning mechanisms on the developmental dynamics of language acquisition. The initial results reported here show that this is indeed possible.

Unavoidably, there are a number of scaffolds and simplifications, which we hope to remove in future work. Specifically, the learning operators used so far have been quite simple, which has required that more constraints than desirable had to be imposed for learning to take off. For instance, in the experiments reported here, agents are biased to build categories on single feature channels because this leads them in a quite straightforward way to proximal, absolute and projective strategies. In future work, we want to avoid this and let agents learn themselves discover that this bias is appropriate.

Another goal of our future work is going to be scaling. Locative spatial language is only a small albeit important aspect of language. There has already been some work within the same paradigm on learning color lexicons [Bleys *et al.*, 2009], the emergence of quantifiers [Pauw and Hilferty, 2012] and parts of tense and aspect systems [Gerasymova and Spranger, 2010]. However, most of these experiments focus on one specific aspect of language. It remains to be studied how a complete (or at least more complete) language can arise from a developmental point of view.

References

[Alishahi and Stevenson, 2008] A. Alishahi and S. Stevenson. A computational model of early argument structure acquisition. *Cognitive science*, 32(5):789–834, 2008.

[Asada *et al.*, 2009] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida. Cognitive developmental robotics: A survey. *Autonomous Mental Development, IEEE Transactions on*, 1(1):12–34, 2009.

[Bailey *et al.*, 1997] D. Bailey, J. Feldman, and S. Narayanan. Modeling embodied lexical development. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society: August 7-10, 1997, Stanford University*, page 19. Lawrence Erlbaum Associates, 1997.

[Bleys *et al.*, 2009] J. Bleys, M. Loetzsch, M. Spranger, and L. Steels. The Grounded Color Naming Game. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-man 2009)*, 2009.

[Brown and Levinson, 1993] P. Brown and S.C. Levinson. "uphill" and "downhill" in tzeltal. *Journal of Linguistic Anthropology*, 3(1):46–74, 1993.

[Brown, 2008] P Brown. Up, down, and across the land: landscape terms, place names, and spatial language in tzeltal. *Language Sciences*, 30(2-3):151–181, 2008.

[Chen and Mooney, 2010] David L Chen and Raymond J Mooney. Learning to interpret natural language navigation instructions from observations. *Journal of Artificial Intelligence Research*, 37:397–435, 2010.

[Frank *et al.*, 2008] Michael Frank, Noah D Goodman, and Joshua B Tenenbaum. A bayesian framework for cross-situational word-learning. In J.C. Platt, D. Koller, Y. Singer, and S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 457–464. Curran Associates, Inc., 2008.

[Gerasymova and Spranger, 2010] K. Gerasymova and M. Spranger. Acquisition of grammar in autonomous artificial systems. In H. Coelho, R. Studer, and M. Woolridge, editors, *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI-2010)*, pages 923–928. IOS Press, 2010.

[Griffiths *et al.*, 2010] T Griffiths, N Chater, C Kemp, A Perfors, and J Tenenbaum. Probabilistic models of cognition: exploring representations and inductive biases. *Trends in cognitive sciences*, 14(8):357–364, 2010.

[Haddock, 1989] N. J. Haddock. Computational models of incremental semantic interpretation. *Language and Cognitive Processes*, 4(3):337–368, 1989.

[Herskovits, 1986] A. Herskovits. *Language and spatial cognition*. Studies in Natural Language Processing. Cambridge University Press, 1986.

[Levinson, 2003] S. C. Levinson. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press, 2003.

[Matuszek *et al.*, 2010] C. Matuszek, D. Fox, and K. Koscher. Following directions using statistical machine translation. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 251–258. IEEE Press, 2010.

[Pauw and Hilferty, 2012] S. Pauw and J. Hilferty. The emergence of quantifiers. In L. Steels, editor, *Experiments in Cultural Language Evolution*. John Benjamins, 2012.

[Regier, 1996] T. Regier. *The human semantic potential: Spatial language and constrained connectionism*. The MIT Press, 1996.

[Retz-Schmidt, 1988] G. Retz-Schmidt. Various views on spatial prepositions. *AI magazine*, 9(2):95–105, 1988.

[Saunders *et al.*, 2009] J. Saunders, C. Lyon, F. Forster, C.L. Nehaniv, and K. Dautenhahn. A constructivist approach to robot language learning via simulated babbling and holophrase extraction. In *Artificial Life, 2009 (ALife'09). IEEE Symposium on*, pages 13–20. IEEE, 2009.

[Saunders *et al.*, 2011] J. Saunders, H. Lehmann, Y. Sato, and C. Nehaniv. Towards using prosody to scaffold lexical meaning in robots. In *Development and Learning (ICDL), 2011 IEEE International Conference on*, volume 2, pages 1–7. IEEE, 2011.

[Siskind, 1996] J. M. Siskind. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1):39–91, 1996.

[Spranger *et al.*, 2012] M. Spranger, S. Pauw, M. Loetzsch, and L. Steels. Open-ended Procedural Semantics. In L. Steels and M. Hild, editors, *Language Grounding in Robots*, pages 153–172. Springer, 2012.

[Spranger, 2013] M. Spranger. Grounded lexicon acquisition - case studies in spatial language. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2013 Joint IEEE International Conferences on*, pages 1–6. IEEE, 2013.

[Steels, 2001] L. Steels. Language games for autonomous robots. *IEEE Intelligent systems*, pages 16–22, 2001.

[Svorou, 1994] S. Svorou. *The Grammar of Space*, volume 25 of *Typologogical Studies in Language*. John Benjamins, 1994.

[Tellex *et al.*, 2013] Stefanie Tellex, Pratiksha Thaker, Joshua Joseph, and Nicholas Roy. Learning perceptually grounded word meanings from unaligned parallel data. *Machine Learning*, 2013.

[Tenbrink and Moratz, 2003] T. Tenbrink and R. Moratz. Group-based spatial reference in linguistic human-robot interaction. In *Proceedings of EuroCogSci'03, The European Cognitive Science Conference 2003*, pages 325–330. Lawrence Erlbaum, 2003.

[Winograd, 1971] T. Winograd. *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language.* PhD thesis, Massachusetts Institute of Technology, 1971.