Acquisition of Grammar in Autonomous Artificial Systems

Kateryna Gerasymova and Michael Spranger[1]

Abstract. Over the past several decades, psycholinguists have gained countless insights into the process of child language acquisition. Can these findings be used for the development of language competence in autonomous artificial systems? This paper reports on our attempt to apply insights from developmental psychology in order to enable artificial systems to acquire language. We consider a comprehensive chain of computational processes, starting from conceptualization and extending through language generation and interpretation, and show how they can be intertwined to allow for acquisition of complex aspects of grammar.

1 INTRODUCTION

Over the last few decades numerous psycholinguistic studies have been concerned with the process of child language acquisition [17, 6]. Many of these studies have attempted to elucidate how the learning process is structured, that is, they have tried to identify strategies and stages which learners undergo in their development of linguistic competence. The most prominent *usage-based* approach to acquisition supports the view that a child's linguistic skills result from her accumulated experience with language and proposes the following typology of children's early utterances: children begin speaking using *holophrastic* units, later they learn *item-based constructions*, and only at the final stage do they manage adult-like *abstract constructions* [17]. Researchers interested in cognitive modeling may therefore wish to investigate whether these findings can be operationalized in a systematic way, and whether they can aid the design of autonomously learning systems.

This paper aims to computationally model processes involved in language acquisition for autonomous artificial agents, motivated by attempts to make concrete instantiations for theories proposed by psychologists and psycholinguists and check their applicability to artificial systems. We especially focus on *online*, *unsupervised* learning because the ability to autonomously learn from limited data and dynamically adapt to the environment and behavior of others may be the most crucial feature of an intelligent cognitive system.

Our investigation makes use of *language games* [12], a framework that has proven fruitful for repeated focused experiments on communication. A language game is a situated embodied interaction between two agents, possibly two robots or a robot and a human. Its situated setting strongly constrains the set of possible interpretations, which makes communication more predictable and provides a framework for semantic inference needed to make verbal communication feasible for the AI endeavor.

This paper focuses on the acquisition of grammar. The core of the paper describes a learning experiment in which artificial agents acquire rules of the target grammar. As a case study we apply the usage-based theory of acquisition to the learning of the grammatical category of Russian aspect. We will especially focus attention on parallels between the presented learning stages and the development of linguistic capabilities in children. However, it is not our explicit aim to model a child acquiring her first language; rather, we draw inspiration from solutions found in nature in order to capture the general learning mechanisms that are essential for open-ended learning and adaptation of autonomous systems. More precisely, the experiment demonstrates computationally how an aspectual grammar can be acquired, through the use of analogies across linguistic constructions, as well as the definition of functional roles based on those analogies.

Before delving into the actual acquisition experiment, we briefly describe in the next section some necessary linguistic concepts and summarize the insights from developmental psychology relevant to the study.

2 THEORETICAL BACKGROUND

2.1 Grammatical category of aspect

Aspect indicates how events unfold in time. For instance in English, the difference between *it rained* and *it was raining* is not one of tense but rather one of aspect, since both refer to the same raining event that occurred in the past [1].

In Russian, aspectual distinctions are ubiquitous, and every verb in all forms is aspectually marked. There are two aspects that have to be distinguished: verbs can be *perfective* or *imperfective*. Imperfective verbs are usually morphologically 'simple,' consisting only of a stem and an ending. They are durative in nature and are often used as a description of usual facts, simple activities and states, for example, говорить (*govorit'*, 'talk').

In contrast, perfective verbs do not express any duration. The definition of the perfective aspect is that it expresses an action as a total event summed up with reference to a single juncture [2]. This means that a particular portion of an action is viewed as a complete event on its own; for example, the formзаговорить (*zagovorit'*, 'begin to talk') expresses the beginning of talking as a complete event. Which portion of an event is highlighted depends on the temporal semantics of the verb, also called *Aktionsart* of the verb. For instance, the *ingressive* Aktionsart refers to a focus on the beginning of the event (as in заговорить), while the terminative Aktionsart refers to the final portion of an event; the *delimitative* Aktionsart denotes the development of an action to a limited extent, whereas the *exhaustive* Aktionsart indicates the performance to an exhaustive degree, etc.

Perfectives are derived from the 'simple' imperfectives by attachment of a prefix (in the example, the prefix за- (*za*-) was attached). And it is the prefix that brings a certain temporal part or a semantic

[1] SONY Computer Science Laboratory Paris, France, email: gera@csl.sony.fr

facet of the event into focus, thereby changing the verb's Aktionsart and aspect. There are nineteen different prefixes that can productively form Perfective [5].

2.2 How children acquire aspect

Children start by acquiring predominantly simple verbs (imperfectives) [3]. This is likely because imperfectives have a minimal stem complexity (simple verbs), do not contain aspectual operators and are easier to use than perfectives. Overall, children acquire aspect in a piecemeal fashion, and learning is not even completed by the age of six [15]. At the beginning, verbal aspect is mastered by children as a part of the lexical meaning of a verb within the general process of cognitive development, i.e., children learn to recognize and to 'name' different situations by means of different forms of verbs. This means that children do not learn aspect as a separate verbal category per se, and that they have no general semantic representation of the grammatical aspect, but rather rely on the lexical class of individual Aktionsarten [3]. In turn, different Aktionsarten – means to highlight parts of events – are learned independently from each other in a context-specific way [15]. Only later is the grammatical category of aspect abstracted away from this tight contextual connection by unifying several Aktionsarten into the abstract category of perfective. At this stage aspect finally becomes recognized as a category separate from the lexical meaning of verbs [3].

We can conclude from these psycholinguistic studies that the acquisition of Russian aspect presents a significant computational modeling challenge, particularly since brute force, offline statistical algorithms (which, for example, cycle through a database of examples and counterexamples annotated in advance) are unrealistic in this setting. Recall that we wish to study how aspect learning can take place within the context of ongoing communicative interactions, which requires one-shot learning from few examples. Such algorithms, however, are rare in the machine learning community.

3 EXPERIMENT

The aim of the current experiment is to show how autonomous agents can acquire the Russian aspectual system from observing language pieces in communication. We build on extensive previous research on lexicon formation [9] and assume that the learning agents are equipped with a fully developed lexicon. The setup of the experiment is inspired by the comprehension experiment of Stoll [15], who investigated how children develop their understanding of aspectual forms. Preschool children were interviewed after watching pairs of short movies, each illustrating what would be described by a different aspectual form of the same verb stem. Similarly, in our experiment artificial agents observe pairs of events differing in temporal semantics and consequently best described by different aspectual forms. Some agents in the population are tutors and possess a fully developed aspectual system, that is, they have a set of grammatical constructions that allows them to successfully process aspectually marked utterances. Further, there are learning agents which have to autonomously acquire the aspectual system. Agents of both types engage in dialogues, and the learning agents subsequently pick up the aspectual grammar, so that at the end of the experiment all agents have converged on the same set of grammatical constructions.

3.1 Language game for aspect

The experiment is based on language games [12], which are routinized communicative interactions between pairs of agents. One possible interaction is as follows:

1. Two agents are randomly selected from the population. One agent acts as a speaker, another one is a hearer.
2. Both agents perceive a shared context, which models their joint attention [16]. The context consists of two events of the same kind but with different temporal semantics, e.g., *ongoing reading* versus *reading for a while*. The two events feature different protagonists (either Миша (Michael) or Маша (Masha)); thus, two example events are *Michael reading for a while* versus *Masha reading the whole time*.
3. The speaker starts the interaction by choosing one event from the context as a *topic*, for example, the event where Michael was *reading for a while*.
4. The communicative goal of the speaker is to ask a question about the protagonist of the topic-event (in our case Michael) using the information about the event he was involved in. The question should unambiguously discriminate the protagonist, which means here that the event's temporal structure has to be incorporated into the question. For example, Кто почитал? (*Kto počital?*, 'Who read for a while?') discriminates Michael because only he was involved in the action for a short period of time (Masha was reading for the whole time). The speaker conceptualizes the meaning of the question and then produces an utterance and transmits it to the hearer.
5. The hearer perceives the utterance, parses it, and then interprets the parsed meaning by comparing the result of interpretation to the context. The task of the hearer is to identify the protagonist of the topic-event unambiguously; guessing is not allowed.
6. If the hearer is able to unambiguously answer the question, she verbalizes her answer by saying *Michael*. Otherwise, he gives up.
7. The speaker signals whether the answer is correct, i.e., whether the answer corresponds to the protagonist of the topic-event. The right answer means communicative success, no answer or a wrong answer is considered to be a communicative failure.
8. In the case of either incorrect or absent answer, the speaker reveals the desired answer.
9. Based on the outcome of the interaction, the learner consolidates his grammatical knowledge by increasing or decreasing the scores of grammatical constructions, as well as creating new constructions or deleting the old ones.

To successfully participate in such an interaction, agents have to manage numerous processes. Acting as a speaker, they must conceptualize the perceived scene, find a discriminative meaning for the chosen topic, build the semantic structure for the potential utterance, as well as produce an utterance to express the intended meaning. Similarly, the listener has to be able to parse utterances and interpret their meaning in order to conceptualize answers to questions or decide on the correctness of given answers. Finally, the learners need mechanisms for invention and adaptation of grammatical knowledge.

3.2 Realization of semantics

The conceptualization process of agents is implemented in *Incremental Recruitment Language* or *IRL* [8], which is a formalism for representing and computing semantic structures that constitute the meaning of utterances. IRL is tightly integrated with the engine for language production and parsing. Its main purpose is to provide semantic structure that is translated into syntactic structure by the language system and back. Consequently, IRL not only encompasses how to represent the semantic structure of an utterance, how to interpret a

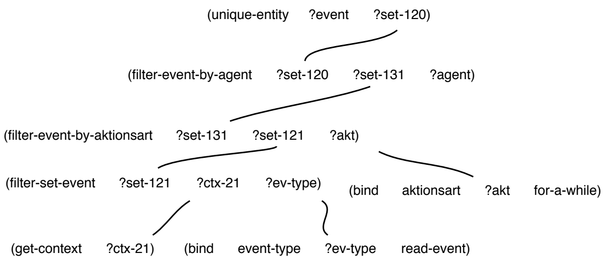

Figure 1. Semantic structure underlying the question Кто почитал? (*Kto počital?*, 'Who read for a while?'). This structure was autonomously constructed by an agent based on the current world model and the communicative goal of asking a question about a particular event.

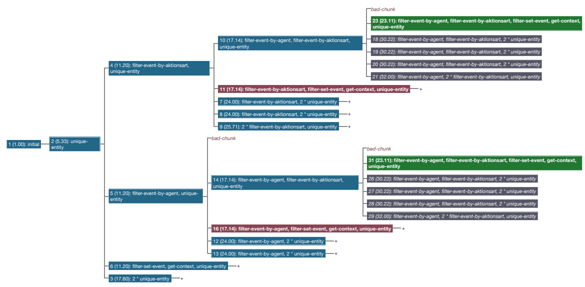

Figure 2. Search tree for constructing the semantic structure from Figure 1. Semantic structure is autonomously created by the agents in order to fulfill communicative goals; in our case, to construct a discriminative question. Each node in this search tree corresponds to an IRL-network. Successful nodes are highlighted in green. Other nodes are unsuccessful with regards to the particular goal and context faced by the agent.

semantic structure given particular communicative contexts, but also how to build a semantic structure autonomously given communicative goals and/or perceived utterances.

In the example interaction of the aspect language game, the communicative goal of the speaker is to formulate a question that discriminates the protagonist of the topic-event from the context. Hence, the semantics of the question has to convey a sequence of mental operations that, when executed, lead to the topic-event and thereby to its protagonist. This sequence of operations, or the semantic structure of the question, is represented as an *IRL-network*. For instance, Figure 1 depicts the IRL-network for the semantic structure underlying the question Кто почитал? (*Kto počital?*, 'Who read for a while?'). IRL-networks link concrete *operations*, e.g., `filter-set-event`, and *entities*, e.g., `read-event`. The latter are introduced using so-called *bind statements* and represent event categories, prototypes, concepts or particular temporal characteristics of events, i.e., Aktionsarten. Operations and semantic entities are linked via variables (whose names start with `?`). The network in Figure 1 reveals that the corresponding question prescribes operations that filter and profile events using, for instance, event prototypes and Aktionsarten.

Upon construction, semantic structures are immediately tried out on the context to examine their discriminative power. When the network from Figure 1 is executed, `get-context` first introduces the shared context into the network and *binds* it to the variable `?ctx-21`. This is followed by the execution of the `filter-set-event` operation that filters all events in the context for those of type `read-event` (introduced via the corresponding bind statement). The output set is bound to the variable `?set-121`, which serves as an input to the `filter-event-by-aktionsart` operation. This operation brings a certain portion of an event (from the input set) into focus, taking into account the relation of this event to the time frame of the observed scene. For instance, an event where neither a start nor an end point was perceived can be filtered as ongoing. This operation is followed by the `filter-event-by-agent` primitive which filters events for those with a particular protagonist. With an unbound agent, this primitive returns all possible event-agent pairs. In the end, the events computed by this network are checked for uniqueness – a test the speaker performs to assure the unambiguous discrimination of one event in the context. The protagonist of the event is bound to the variable `?agent`.

Semantic structure, like the one just discussed, is autonomously constructed by agents either when producing an utterance or when interpreting one. Agents are equipped with a set of mental operations and entities, and they construct the IRL-networks in a heuristically guided search process (see Fig. 2), where more and more operations are added to the network. The networks are then executed against the context, whereby it is checked whether the particular communicative goal of the agent is met, i.e., for the speaker to construct a question about a particular event in the environment or for the hearer to interpret a partial structure parsed by the language system.

3.3 Language processing

Language processing is implemented in the *Fluid Construction Grammar* (*FCG*) formalism [13], which is a computational implementation of Construction Grammar that is integrated with IRL. IRL networks constitute the semantic input for the FCG grammar engine. In line with cognitive linguistics, FCG represents all linguistic knowledge in a uniform way, namely as constructions that map some aspects of meaning onto some aspects of form and back [4]. Constructions are represented as *coupled feature structures*, coupling semantic and syntactic poles. Both semantic and syntactic poles consist of units, which, in turn, are organized in feature-value pairs. The units might correspond to lexical items, morphemes, words, or phrases. Semantic features of a unit are related to semantic aspects, for example, the associated meaning or the semantic categories; syntactic features are responsible for syntactic aspects, for example, the form constraints associated with a unit or its syntactic categories.

Language production and parsing are based on the unify-and-merge principle [14]. Before any grammatical construction can apply to a linguistic structure, it has to satisfy the unification requirement. Only thereafter is the merging of linguistic information performed, enriching the resulting linguistic structure. During language processing, an agent tries to apply all available constructions, building a search tree of possible sequences of applied constructions. In case of alternatives, the branch with the highest estimated success will be chosen. This success score is based on the linguistic context in which constructions are applied and on how successful the applied constructions have been in previous communicative situations.

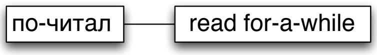

Figure 3. Schema of the *holophrasis* почитал (*počital*, 'read-for-a-while'). This holophrastic construction maps the form of the observed utterance почитал to its meaning *read-for-a-while*. The learner treats it as a single unit without knowing its composition.

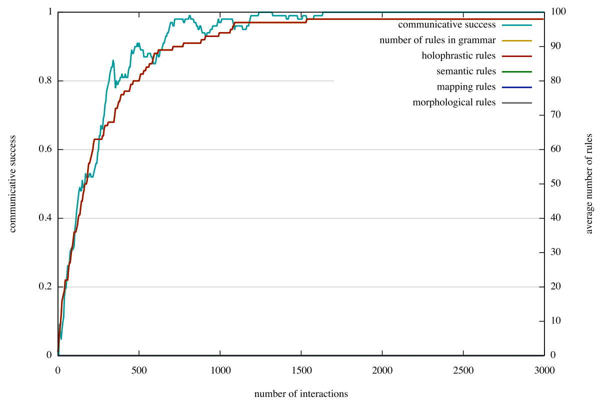

Figure 4. Learning holophrastic constructions. The learner is equipped only with one repair strategy – internalize observed utterances. The communicative success is reached, but the inventory contains 98 holophrastic constructions (14 verbs × 7 different temporal semantics).

3.4 Learning

Through repeated interactions of the aspect language game, the learning agents gradually pick up the aspectual target grammar. The key to agents' ability to learn are cognitive mechanisms for detecting and solving problems that may be encountered during interactions, e.g., inability to parse an utterance or ambiguity in interpretation. Successful application of these problem-solving tactics underlies the whole learning process, which can be divided into (at least) three subsequent stages with respect to the learning mechanisms employed: acquisition of *holophrases*, *item-based constructions* and *abstract constructions*.

3.4.1 Holophrases

Developmentally, *holophrases* are the first type of children's early constructions, where children use a single linguistic symbol to communicate their intentions about a specific scene [17]. Analogically, the learning artificial agents acquire holophrastic constructions during their first phase of learning. This happens when the hearer (learning agent) cannot completely parse a question the speaker posed, as in the example interaction Кто почитал? (*Kto počital?*, 'Who read for a while?'). The linguistic parts that can be processed are *kto* and *čital*, since agents are assumed to have a developed lexicon, but the prefix *po-* is left unprocessed. This leads to ambiguity in the interpretation of the topic-event (since both events are about *reading*), and consequently, two hypotheses about the protagonist involved in the event are created. Since the hearer is not allowed to guess, she gives up, the interaction is a failure, and at the end, the speaker reveals the right answer: *Michael*. The hearer tries to learn from this shortcoming and first stores the complete perceived utterance as a sample. Additionally, she searches her context for a semantic factor that could differentiate *Michael* from *Masha*, since questions are assumed to be discriminative. The distinctive feature for *Michael* is the temporal structure of his *reading*, which is *for a while*, in contrast to the *ongoing reading* of *Masha*. The stored sample is supplemented with this deduced information (schematically shown in Figure 3). The holophrasis is implemented as an FCG construction – mapping of meaning and form. *Kto* is not stored in the sample construction because it is assumed to be known by the agent.

The intuition behind holophrases is that the learning agent assumes that почитал is a single constituent after encountering it for the very first time. This way, the learning agent stores perceived samples creating undifferentiated holophrastic constructions, e.g., поиграл (*poigral*, 'played-for-a-while'), порисовал (*porisoval*, 'drew-for-a-while'). These holophrasis constructions are fully operational, which means that by the second time the agent hears the same question, she will be able to parse it entirely and, moreover, generate this question when in the role of the speaker (but only the exact same question).

When learners are equipped with such an internalization strategy, they are able to communicate successfully after memorizing all possible prefix+verb combinations they have encountered. Figure 4 depicts the convergence of communicative success accompanying subsequent acquisition of holophrases. However, such organization of the language inventory is unsatisfactory. With every additional verb floating in the population, the number of needed constructions increases by the number of temporal semantic features, and with every additional semantic feature by the number of verbs. Furthermore, such inventory organization lacks any notion of grammar, which contradicts the known abilities of adult native speakers of Russian to recognize two distinct aspects.

In summary, this stage closely resembles how children initially master verbal aspect as part of the lexical meaning of the verb.

3.4.2 Item-based Constructions

In the second stage, the learning agent creates more general constructions based on repeatedly encountered samples of similar kind; developmentally this corresponds to the emergence of *item-based constructions* [17]. For example, the hearer again faces the problem of ambiguous interpretation because of the inability to parse пописал (*popisal*, 'wrote-for-a-while'). But now, instead of giving up, she searches through her stored samples for a means of parsing the utterance, eventually noticing that the difference between holophrastic constructions for почитал, поиграл, порисовал is the actual verb stem. Hence, the agent is able to create a more general construction for the usage pattern по+*verb* (with a slot for a verb), as shown in Figure 5, and successfully parse the utterance involving пописал. The discovery of this usage pattern corresponds to the acquisition of the delimitative Aktionsart. More precisely, the agent has learned that the presence of the prefix по- (*po-*) in front of a verb indicates that the temporal semantic feature *for a while* has been added to its meaning.

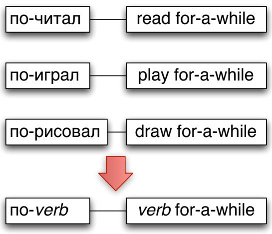

Figure 5. Learning of the item-based construction по+verb (*po+verb*, 'verb+for-a-while'). Above the arrow: undifferentiated holophrases are stored in memory when encountered. Under the arrow: holophrases with a particular prefix become generalized to an item-based construction based on this prefix, enabling parsing of prefixed verbs.

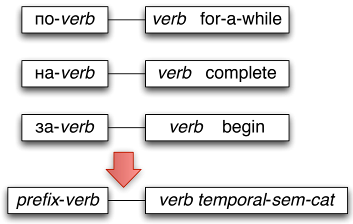

Figure 6. Learning principle of prefixation: emergence of an abstract construction for perfective. Above the arrow: item-based constructions based on particular prefixes. Under the arrow: abstract construction expressing the general principle of prefixation for derivation of new Aktionsarten is learned through generalization over item-based constructions.

After this stage, the learning agent can correctly interpret any (known) verb prefixed by по- (*po-*), even if she has not encountered this particular combination before. However, the agent has acquired *only* the ability to comprehend the pattern по+*verb* but is still lacking additional knowledge for utilizing this device in language generation. What is missing is an understanding of the general principle of deriving new Aktionsarten by prefixation required to actively create a prefix structure in production.

Nevertheless, the process of generalization described here for the prefix по- (*po-*) works exactly the same for other prefixes, given enough generalizable material in an agent's linguistic inventory. The learned constructions are item-based, the item being the particular prefix. The independent emergence of such item-based constructions for other prefixes mirrors the independent acquisition of Aktionsarten hypothesized in Section 2.2.

3.4.3 Abstract Constructions

The final phase of the acquisition process in children is characterized by generalization over item-based constructions and formation of *abstract constructions*, in which children express their communicative intentions through utterances that instantiate relatively abstract, adult-like linguistic constructions [17].

Although the two previous learning strategies solved the parsing problem for artificial learners, production remains troublesome. When faced with a need to generate a question in dialogs, learners are still unable to construct the complete utterance. In particular, they are unable to express the temporal semantics of events needed for discrimination. This failure is detected by the learner after *re-entering* the outcome of production into her own language system for parsing and noticing that the constructed utterance is insufficient to single out the topic. The idea behind re-entrance is to predict the effect of the utterance before actually passing it to the hearer.

To repair her communicative problem, the learner examines the inventory of her linguistic experiences. There, accumulated item-based constructions reveal a general principle that the temporal semantics of verbs (Aktionsart) can be expressed by means of prefixation. This discovery can be captured by a novel abstract construction, where a prefixed verb, regardless of the actual form of the prefix and corresponding Aktionsart, becomes marked for the perfective aspect. The new construction operates only on the abstract semantic and syntactic categories of Aktionsart (*temporal-sem-cat*) and aspect and generates an abstract unit for a prefix without any concrete linguistic material (Figure 6). Only after this stage is the agent able to generate the perfective derivation of any (known) verb without having heard the resulting form before.

The described process resembles the way children acquire the grammatical category of aspect late in development, by unifying several Aktionsarten into the abstract category of perfective.

4 RESULTS

Through repeated interactions of the presented aspect language game, artificial learners are able to acquire the aspectual grammar. Figure 7 (upper graph) displays the development of the grammar of one learning agent. In the world in which the learner is situated, events can exhibit 7 different temporal semantics: *ongoing, begin, for a while, finish, complete, exhaustion, alteration*. Therefore, the target grammar should contain 20 construction in total.[2] In the beginning, the only kind of grammatical constructions the learning agent creates are holophrases (red line); their number is aligned with the total number of grammatical constructions the agent acquires (yellow line). After a couple dozen interactions, the learner starts to generalize, noticing the system behind the stored samples: other types of grammatical constructions are generated (semantic and morphological item-based and abstract mapping constructions, indicated by the green, gray and blue lines, respectively). The communicative success converges to the maximum value after approximately 300 interactions (cyan line); each posterior game will be a success.

All constructions in the agent's inventory have a score in the range of [0..1] at any given time during the game. When a new rule comes into play, it is assigned an initial score of 0.5. In the course of the game, the scores of constructions are updated depending on their success in communication (unsuccessful constructions are punished). After the target grammar is acquired (20 constructions in total), the very specific holophrastic constructions become redundant: they are in competition with more general item-based and abstract constructions. Eventually, holophrases lose and disappear after about 2000

[2] This number results from the particular realization of the target grammar in FCG and is assembled from 7 semantic and 7 abstract mapping constructions (for each temporal semantic facet) and 6 morphological constructions (durative Aktionsart coding the *ongoing* temporal semantics does not require a prefix and, therefore, lacks a morphological construction).

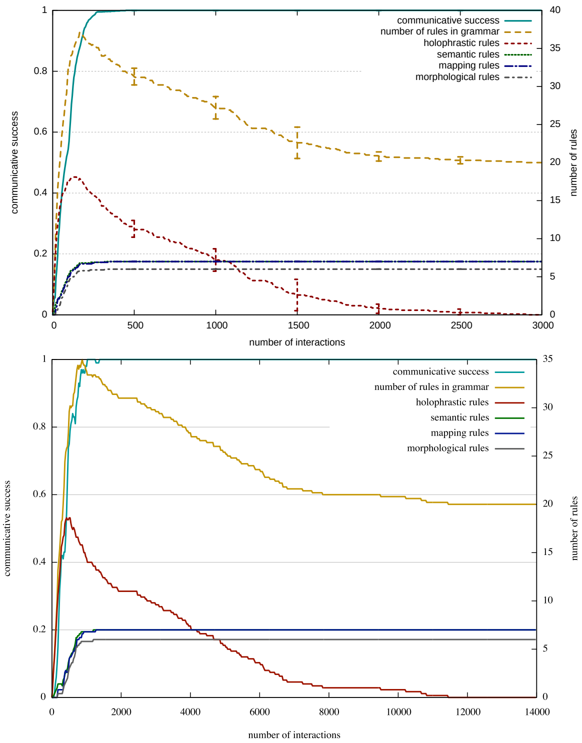

Figure 7. Development of aspectual grammar: communicative success and number of grammatical constructions of one learner during the acquisition process (above: population of one learner and one tutor, avg. of 10 parallel runs of the experiment; bottom: population of 10 agents with 5 learners).

interactions. The bottom graph in Figure 7 displays a similar dynamic for the scaled-up case of 5 learning agents in a population of 10 agents.

5 DISCUSSION

To reconstruct the presented results on real robots, one has to consider the issue of grounding [7, 11], which in the context of this paper is best understood as the question of how operations like `filter-set-event` and semantic entities like `read-event` relate to the real world. Although there has been much work on the grounding of event descriptions [10], it remains to be seen how abstract items like Aktionsarten and their use in filter operations can be grounded in sensorimotor interaction with the environment. However, it is important to realize that whatever the concrete implementation might be, it will most probably not affect the results of this paper. This is due to the open nature of IRL, which makes no claims about the concrete implementation of the operations in networks.

Despite the grounding issue, the results in this paper show that agents are able to acquire an aspectual system starting from barely more than a lexicon. These results therefore support a developmental route to linguistic competence without the necessity of an elaborate innate linguistic structure. Moreover, the results also hint at a developmental route for the development of artificial cognitive systems, as they show how agents can self-organize and learn a language system based only on interactions with peers and equipped with the suitable set of learning operators. Since these learning operators stem from investigations into general cognitive development in children, artificial systems equipped with these learning operators will most likely be able to use the same set of generalization mechanisms for acquiring other complex grammatical systems, such as, for example, case.

Finally, the usage-based approach to acquisition and construction grammar turn out to be both a plausible and a solid basis for modeling the development of linguistic competence in artificial systems.

6 CONCLUSION

This paper investigated how artificial learners can acquire grammar through communicative interactions. Artificial tutors equipped with subsets of Russian aspectual grammar and learning agents interacted, forcing students to infer and adopt constructions for talking about aspectually marked events in their environment. We introduced three different learning operators reminiscent of findings in developmental psychology. Results proved that the proposed learning operators together with the machinery needed for routine conceptualization and language processing can lead to the successful adoption of the structure present in the grammar.

REFERENCES

[1] Bernard Comrie, *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*, Cambridge: Cambridge University Press, 1976.

[2] James Forsyth, *A Grammar of Aspect: Usage and Meaning in the Russian Verb*, Cambridge: Cambridge University Press, 1970.

[3] Natalia Gagarina, 'The acquisition of aspectuality by Russian children: the early stages', *ZAS Papers in Linguistics*, **15**, 232–246, (2000).

[4] Adele E. Goldberg, *A Construction Grammar Approach to Argument Structure*, Chicago UP, Chicago, 1995.

[5] Maksim A. Krongauz, *Pristavki i glagoly v russkom jazyke: semantičeskaja grammatika*, Moscow: Jazyki russkoj kul'tury, 1998.

[6] E. Lieven, H. Behrens, J. Speares, and M. Tomasello, 'Early syntactic creativity: A usage-based approach', *Journal of Child Language*, **30**(02), 333–370, (2003).

[7] D. Roy, 'Grounding words in perception and action: computational insights', *Trends in Cognitive Sciences*, **9**(8), 389–396, (2005).

[8] Michael Spranger, Martin Loetzsch, and Simon Pauw, 'Open-ended grounded semantics', in *Proceedings of 19th Europeen Conference on Artificial Intelligence (ECAI 2010)*, (2010).

[9] L. Steels, 'A self-organizing spatial vocabulary', *Artificial Life*, **2**(3), 319–332, (1995).

[10] L. Steels and J.C. Baillie, 'Shared grounding of event descriptions by autonomous robots', *Robotics and Autonomous Systems*, **43**(2-3), 163–173, (2003).

[11] L. Steels and P. Vogt, 'Grounding adaptive language games in robotic agents', in *Proceedings of the fourth european conference on artificial life*, pp. 474–482. The MIT Press, (1997).

[12] Luc Steels, 'Language games for autonomous robots', *IEEE Intelligent systems*, 16–22, (2001).

[13] Luc Steels, 'Constructivist development of grounded construction grammars', in *Proceedings 42nd Annual Meeting of the Association for Computational Linguistics*, ed., Walter Daelemans, pp. 9–19, Barcelona, (2004). ACL.

[14] Luc Steels and Joachim De Beule, 'Unify and merge in Fluid Construction Grammar', in *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, eds., P. Vogt, Y. Sugita, E. Tuci, and C. Nehaniv, LNAI 4211, pp. 197–223, Berlin, (2006). Springer.

[15] Sabine Stoll, 'The role of aktionsart in the acquisition of Russian aspect', *First Language*, **18**(54), 351–376, (Oct 1998).

[16] Michael Tomasello, *Joint attention as social cognition*, Erlbaum, Hillsdale, NJ, 1995.

[17] Michael Tomasello, 'First steps toward a usage-based theory of language acquisition', *Cognitive Linguistics*, **11-1/2**, 61–82, (2000).