Modelling Language Acquisition through Syntactico-Semantic Pattern Finding

Jonas Doumen[1,2] and **Katrien Beuls**[3] and **Paul Van Eecke**[1,2,4]

[1]KU Leuven, Faculty of Arts, Research Unit Linguistics, Blijde Inkomststraat 21, B-3000 Leuven
[2]KU Leuven, imec research group itec, Etienne Sabbelaan 51, B-8500 Kortrijk
[3]Faculté d'informatique, Université de Namur, rue Grandgagnage 21, B-5000 Namur
[4]Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels

`jonas.doumen@kuleuven.be`
`katrien.beuls@unamur.be`
`paul@ai.vub.ac.be`

Abstract

Usage-based theories of language acquisition have extensively documented the processes by which children acquire language through communicative interaction. Notably, Tomasello (2003) distinguishes two main cognitive capacities that underlie human language acquisition: intention reading and pattern finding. Intention reading is the process by which children try to continuously reconstruct the intended meaning of their interlocutors. Pattern finding refers to the process that allows them to distil linguistic schemata from multiple communicative interactions. Even though the fields of cognitive science and psycholinguistics have studied these processes in depth, no faithful computational operationalisations of these mechanisms through which children learn language exist to date. The research on which we report in this paper aims to fill part of this void by introducing a computational operationalisation of syntactico-semantic pattern finding. Concretely, we present a methodology for learning grammars based on similarities and differences in the form and meaning of linguistic observations alone. Our methodology is able to learn compositional lexical and item-based constructions of variable extent and degree of abstraction, along with a network of emergent syntactic categories. We evaluate our methodology on the CLEVR benchmark dataset and show that the methodology allows for fast, incremental and effective learning. The constructions and categorial network that result from the learning process are fully transparent and bidirectional, facilitating both language comprehension and production. Theoretically, our model provides computational evidence for the learnability of usage-based constructionist theories of language acquisition. Practically, the techniques that we present facilitate the learning of computationally tractable, usage-based construction grammars, which are applicable for natural language understanding and production tasks.

1 Introduction

Usage-based theories of language acquisition argue that the ability of children to learn language is based on two general cognitive capacities: *intention reading* and *pattern finding* (Tomasello, 2003, 2009). Intention reading refers to the capacity of children to understand the communicative intentions of their interlocutors. Pattern finding refers to the ability to recognise similarities and differences in sensory-motor experiences, and to use this ability for categorisation and schema formation (Tomasello, 2003, p. 3–4). Pattern finding thus provides mechanisms for generalising across different communicative interactions, thereby constructing abstract schemata that represent the linguistic knowledge of a language user. In the context of language acquisition, intention reading and pattern finding are two key cognitive capacities that are highly complementary. Intention reading allows a language learner to reconstruct the meaning of an utterance that they observe during a communicative interaction. Pattern finding then provides the mechanisms to learn a grammar based on the combination of these observed utterances and their reconstructed meanings.

There exists an impressive body of theoretical and empirical evidence for both intention reading (Bruner, 1983; Sperber and Wilson, 1986; Meltzoff, 1995; Nelson, 1998) and pattern finding (Goldberg, 1995; Croft, 2000; Diessel, 2004; Goldberg, 2006). However, no comprehensive mechanistic models that provide a faithful operationalisation of either of these cognitive processes exist to date. In this paper, we aim to fill part of this void by presenting a computational operationalisation of pattern finding mechanisms that can bootstrap a grammar based on a set of semantically annotated utterances alone. As such, we assume that the outcome of the intention reading process is given, hence the availability

1347

of the utterances' semantic representations, but that neither a segmentation of the utterances nor any pre-existing morpho-syntactic or other grammatical information can be used. For a computational model that operationalises the intention reading process, and that integrates it with the pattern finding mechanisms introduced in this paper, we refer the interested reader to Nevens et al. (2022).

A validation of our methodology on the CLEVR benchmark dataset for visual question answering (Johnson et al., 2017) shows that it allows for fast, incremental and effective grammar learning. The result of this learning process is a fully-operational, productive construction grammar that can be used for both language comprehension, i.e. mapping from utterances to their meaning representation, and language production, i.e. mapping from a meaning representation to an utterance.

The scientific contribution of this paper is twofold. On the one hand, it provides computational evidence for the cognitive plausibility of usage-based theories of language acquisition by introducing a mechanistic model of the acquisition of construction grammars from scratch. On the other hand, the techniques that we present pave the way for learning computationally tractable, large-scale, usage-based construction grammars that facilitate both language comprehension and production. Apart from their theoretical importance, such grammars are also highly valuable for a large range of application domains, including intelligent conversational agents (Verheyen et al., 2022) and the semantic analysis of discourse (Willaert et al., 2020; Beuls et al., 2021).

The remainder of this paper is structured as follows. Section 2 presents the dataset, task and learning problem that we address. Section 3 introduces our novel methodology for learning construction grammars. Section 4 presents the evaluation results. Related work is discussed in Section 5. A concluding discussion is provided in Section 6.

2 Data

There are two main requirements for datasets to be compatible with the methodology that we present in this paper. First of all, they need to consist of utterances that are annotated with a representation of their meaning. Second, they need to be large enough so that they contain enough utterances that are similar to each other, but not equal, in terms of either form or meaning. The availability of ex-

emplars that are sufficiently close to each other is a necessary precondition for any generalisation process and is fully consistent with the prevailing hypotheses of how children acquire language (Tomasello, 2003). The exact required size of a dataset is as a consequence directly related to the variety and the degree of complexity of the utterances and meaning representations that it contains.

In this paper, we present and validate our methodology using the CLEVR dataset for visual question answering (Johnson et al., 2017). The utterances in the dataset are semantically annotated and the dataset contains ample examples of utterance-meaning pairs that are similar but not equal to each other. The utterances are English questions about images of scenes depicting different configurations of geometrical figures. Each question is annotated with a semantic representation that captures the logical meaning that underlies it. An example of such a scene, a question and its semantic representation is shown in Figure 1.

The semantic representation in Figure 1 takes the form of a set of predicates that share arguments with each other. In the figure, the predicates are drawn in the form of a network, based on the variables that they share. The meaning representation of a question can naturally be represented as a query, i.e. a series of steps that need to be taken in order to answer the question. Each predicate represents a step in this reasoning process, and intuitively corresponds to an atomic cognitive operation that a human or machine can perform. In the case of the example utterance 'How many rubber spheres are there?', the reasoning process consists of four main steps. The first predicate, GET-CONTEXT, binds the image to the variable '?source'. Then, the FILTER predicate filters the image for instantiations of the concept of SPHERE. The result of this filtering operation, i.e. the set of all spheres that are in the image, is bound to the variable '?spheres'. This set of spheres is subsequently filtered by another FILTER predicate for instantiations of the concept of RUBBER. The resulting set of rubber spheres is bound to the variable '?rubber-spheres'. Finally, the set of rubber spheres is counted by the COUNT predicate and the result is bound to the variable '?nr-of-rubber-spheres'. The meaning of the question 'How many rubber spheres are there?' corresponds thus informally to filtering an image for spheres, filtering the spheres for rubber objects and counting the result of this
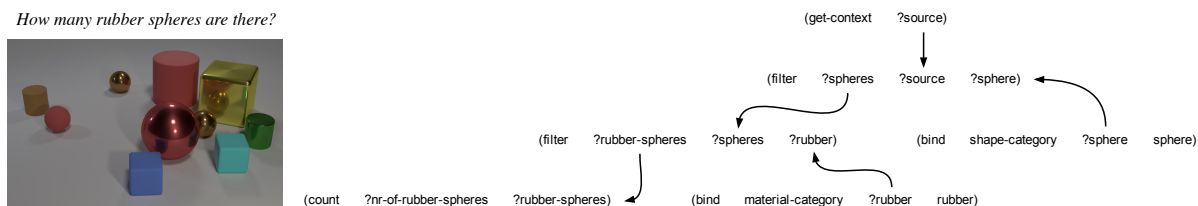
Figure 1: Example scene, question, and procedural semantic representation from the CLEVR dataset.

last operation. Such meaning representations are called *procedural semantic representations* as the representations themselves are at the same time executable procedures (Winograd, 1972; Johnson-Laird, 1977). Our methodology handles procedural semantic representations without problems, but is in no way restricted to it. It can handle any semantic representation, as long as it embraces some notion of compositionality and can be expressed as a set of predicates. Examples of other compatible semantic representations include abstract meaning representation (Banarescu et al., 2013), PropBank frames (Palmer et al., 2005) and the lambda calculus (Church, 1932; Montague, 1974).

The CLEVR dataset consists of three splits: a training split of 70,000 images and 699,989 questions, a validation split of 15,000 images and 149,991 questions, and a test split of 15,000 images and 149,988 questions. The questions in the training and validation splits come with semantic annotations, whereas the test set does not. As we require these annotations in order to evaluate our model, we use the training split of the CLEVR dataset as training set and the validation split as test set. The question-annotation pairs embrace various aspects of reasoning, including attribute identification ('*There is a large cube; what is its color?*'), counting ('*How many green spheres are there?*'), comparison ('*Are there an equal number of large cubes and small things?*'), spatial relationships ('*What size is the cylinder that is right of the yellow shiny thing that is left of the cube?*') and logical operations ('*How many objects are either red cubes or yellow cylinders?*'). For the purposes of this paper, we have selected the subset of CLEVR questions that do not involve comparison, spatial relationships or logical operations. The main reason for this is that these are complex cognitive operations that often correspond to long and complex utterances that are far removed from the linguistic expressions that children (or even other humans) are faced with. Our final training and test sets

consist of 47,134 questions and 10,044 questions respectively.

The learning task that we address consists in operationalising pattern finding mechanisms that facilitate the learning of a bidirectional construction grammar. The grammar should be able to map between the CLEVR utterances and their semantic representation, both in the comprehension (form to meaning representation) and the production (meaning representation to form) direction.

3 Methodology

The input to the learning process consists of utterances that are annotated with a representation of their meaning. The output of the learning process should consist in form-meaning mappings (constructions) that can be used for comprehending and producing utterances. The form-meaning mappings are represented in, and processed using, Fluid Construction Grammar (Steels, 2011; Van Eecke and Beuls, 2017; van Trijp et al., 2022).

3.1 Holophrase Constructions

Let us for a moment take the perspective of the learning algorithm. At the beginning of the learning process, the construction inventory is empty and the first utterance-meaning pair from the corpus comes in. At this point, the only thing that the learning algorithm can do is to store an exact mapping between the observed form and its meaning. Such a holistic mapping corresponds to a holophrase construction and is usable as such, albeit only for comprehending and producing the exact same utterance as the one that was observed. In order to use such a construction in the comprehension direction, it suffices to match the form side of the construction with an utterance and return the meaning side of the construction if the matching process succeeded. In order to use the same construction in the production direction, the meaning side of the construction must be matched with a semantic network and the form side must be returned.

When a next observation comes in, the learning algorithm first checks whether it is already covered by constructions that have been acquired previously. When this is the case, the constructions that are involved in the successful comprehension and production of the observation are reinforced by incrementing their entrenchment score. If the observation is not covered, the algorithm checks whether there are any generalisations that can be made based on the combination of the novel observation and any previously acquired constructions. It is these generalisation mechanisms that embody Tomasello (2003)'s pattern finding capacity and are thereby at the core of the construction learning process. We have identified three classes of mechanisms that facilitate the learning of general constructions by algorithmically reasoning over similarities and differences between existing constructions and novel observations.

3.2 Generalising over Holophrase Constructions

The first class of mechanisms facilitates the generalisation of holophrase constructions with respect to novel observations. These mechanisms can learn item-based constructions that capture the similarities between a novel observation and an existing holophrase construction that was learnt based on a similar, but not equal, observation. These item-based constructions abstract away from the differences between the observation and the holophrase construction.

For example, imagine that a holophrase construction has already been learnt based on the observation of the utterance '*How many rubber spheres are there?*' and the semantic network shown in Figure 1. Now, a novel utterance '*How many rubber cubes are there?*' is observed, along with a very similar meaning network in which the predicate '(bind shape-category ?cube cube)' appears at the place of '(bind shape-category ?sphere sphere)'. The generalisation mechanisms compute the similarities and differences between the construction and the observation in terms of both form and meaning, and make a new item-based construction that maps between the utterance '*How many rubber ?X are there*?' and the semantic network from Figure 1 in which the non-overlapping predicate has been replaced by a variable. At the same time, two new lexical constructions are created, which capture the differences between the observation and the

original holophrase construction. In our example, these will be a construction that maps between the utterance '*cubes*' and the meaning representation '(bind shape-category ?cube cube)' and a construction that maps between the utterance '*spheres*' and the meaning representation '(bind shape-category ?sphere sphere)'. Finally, categorial links are made between the ?X slot in the item-based construction and the new lexical constructions. These categorial links capture that '*cubes*' and '*spheres*' can both appear in the ?X slot of the construction for '*How many rubber ?X are there?*'. A schematic representation of this learning process is shown in Figure 2.

There are three different scenarios in which mechanisms of this class are active. The first scenario concerns utterances which extend holophrases that are already known. An example would be the generalisation of '*Are there any cylinders?*' to '*Are there any red cylinders?*'. In this case, an item-based construction '*Are there any ?X cylinders?*' is learnt, along with a lexical construction for '*red*' and a categorial link between the lexical construction and the open slot in the item-based construction. The second scenario concerns utterances which reduce known holophrases. An example would be the reduction of '*What is the size of the metal block?*' to '*What is the size of the block?*'. In this case, an item-based construction for '*What is the size of the ?X block?*' is learnt, along with a holophrase construction for '*What is the size of the block?*', a lexical construction for '*metal*', and a categorial link between the slot in the item-based construction and the lexical construction. The final scenario concerns utterances which are not a mere extension or reduction of each other, but contain different formal and/or semantic material. An example would be the utterances '*How many rubber spheres are there?*' and '*How many rubber cubes are there?*' discussed above, where a holophrase construction for '*How many rubber spheres are there?*' is already in place. An item-based construction for '*How many rubber ?X are there*' is learnt along with a lexical construction for '*cubes*' and a categorial link between the open slot in the item-based construction and the new lexical construction. Additionally, a second lexical construction for '*spheres*' is learnt, along with a categorial link between the open slot in the item-based construction and the lexical construction for '*spheres*'.
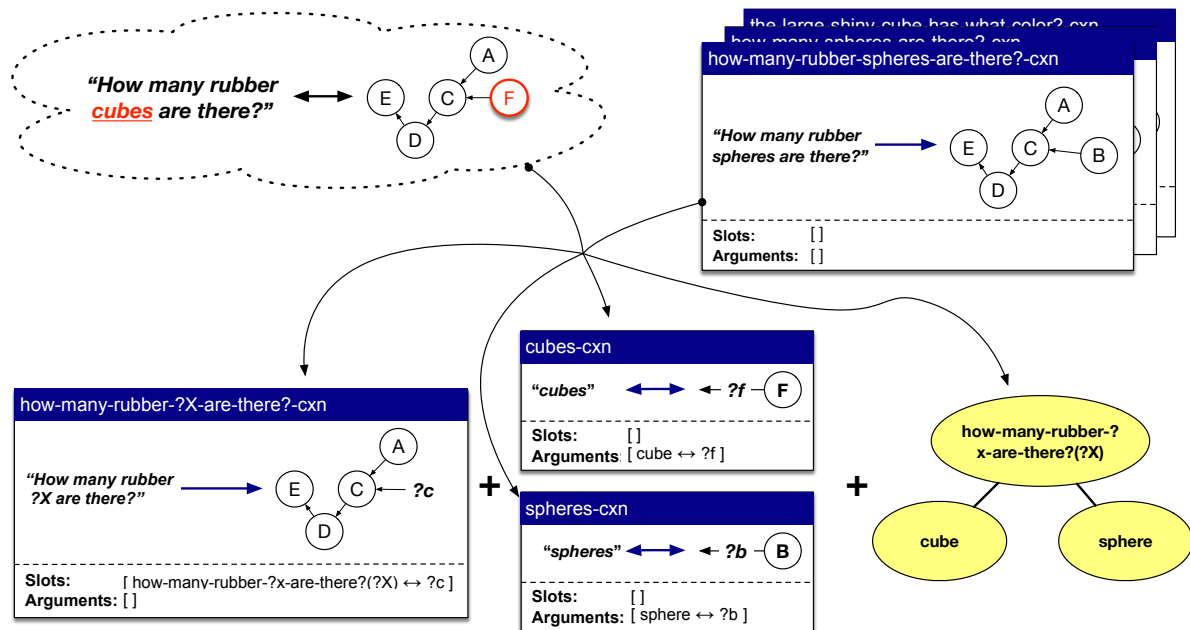
Figure 2: A schematic representation of a generalisation operator learning an item-based construction and two lexical constructions based on an existing holophrase construction and a novel observation.

3.3 Learning Constructions Based on a Partial Analysis

The second class of mechanisms is designed to handle cases where an observation could not completely be processed using the existing constructions of a grammar, but where a partial analysis could be provided. These mechanisms can then create novel constructions that can work together with existing constructions so that the entire observation can be processed successfully. They start thus from the combination of a novel observation on the one hand, and an item-based construction or one or more lexical constructions on the other. The second class of mechanisms is active in two different scenarios.

The first scenario concerns observations to which an item-based construction can apply, but where there remains material that is not covered by any of the existing constructions. An example would be an observation of '*What is the size of the green block?*', where a construction for '*What is the size of the ?X block?*' is already known, while no construction for '*green*' has been learnt yet. The learning algorithm detects that some aspects of the form and the meaning of the observation are not covered by the existing item-based construction and it creates a novel lexical construction that maps between those parts of the form and meaning that were not covered. Additionally, a categorial link is made between the slot in the item-based construc-

tion and the lexical construction. In our example, this means that a lexical construction for '*green*' is learnt, along with a categorial link between this construction and the ?X slot in the construction for '*What is the size of the ?X block?*'.

The second scenario concerns observations to which one or more lexical constructions can apply, but where these constructions do not fully cover the input. An example would be an observation of the utterance '*There is a big red cube; what is its material?*', where lexical constructions for '*big*', '*red*', '*cube*' and '*material*' have already been learnt. The learning algorithm will then create a new item-based construction that incorporates all the form and meaning material that remains after the application of these lexical constructions, and that abstracts away from these constructions through the integration of four slots. The result is an item-based construction of the form '*There is a ?A ?B ?C; what is its ?D?*' and four categorial links from the existing lexical constructions to the slots in the new item-based construction.

3.4 Extending the Categorial Network

The third class of mechanisms is designed to handle cases where all necessary constructions are already in place, but where they cannot combine due to the absence of certain links in the categorial network. An example would be the utterance '*How many things are there?*' where an item-based construc-

tion covering '*How many ?X are there?*' and a lexical construction covering '*things*' already exist, but where there is no link in the categorial network between the lexical construction for '*things*' and the '?X' slot in the item-based construction. In such cases, the learning algorithm adds the missing link to the categorial network.

3.5 Entrenchment Scores

The constructions created by the learning operators have scores that reflect their entrenchment. During processing, higher scored constructions are preferred over lower scored ones. Upon creation, the score of a construction is set to 0.5. If used successfully, the score is increased by 0.1 and the score of other constructions of which the application would also have led to a solution is decreased by 0.3. The scores are bounded between 0 and 1. There is no built-in bias towards more general constructions. However, the fact that more general constructions are applicable in a broader range of situations and are therefore more frequently used, will, due to the dynamics of rewarding successful usage and punishing competitors, lead to higher entrenchment scores for more general constructions.

4 Experiments

This section presents a validation of our methodology for acquiring constructions on the CLEVR dataset discussed in Section 2. We first describe the experimental set-up (Section 4.1) and then present the evaluation results (Section 4.2).

4.1 Experimental Set-Up

The primary experiment consists in processing the 47,134 observations from our training set using the learning operators introduced above. For each experimental run, the observations are shuffled, so that any side-effects that might be caused by the order in which the observations are presented are levelled out. The learning operators are only active when an observation cannot be processed successfully by the constructions that have been learnt so far. Entrenchment scores are updated after each communicative interaction. The learning process is evaluated through four quantitative metrics: *communicative success*, *grammar size*, *number of constructions per type* and *active learning operators*. Communicative success is a binary measure computed by comparing the comprehended meaning with the gold standard annotation. In the graphs

below, communicative success and active learning operators are plotted using a sliding window of 50 observations.

For completeness, we also present a secondary experiment in which the grammar learnt on the training set is evaluated on the test set. Communicative success is here averaged over the whole test set, and grammar size and number of constructions per type do not change during evaluation.

The experimental results reported below are based on 10 independent experimental runs. The error bars that are plotted represent percentiles 5 and 95.

4.2 Results

The results obtained through the primary experiment are shown in Figures 3 to 5. Figure 3 displays the communicative success and grammar size metrics respectively on the left and right y-axis as a function of the number of observations (x-axis). We can see that the communicative success starts at 0, as the experiment starts with an empty inventory of constructions. The degree of communicative success rises rapidly, with more than 90% of the observations being successfully processed by the learned grammar after only 500 observations have been encountered. After 2000 observations, communicative success is already achieved in 99.6% of new observations.

The grammar size starts at 0 constructions and grows rapidly in the first phase of the experiment. After 500 observations, the grammar has reached its peak size of around 230 constructions that have some degree of entrenchment. This number then declines as a result of the rewarding and punishing of constructions. At the end of the learning process, the resulting grammar consists of 101.5 constructions on average.

An analysis of the types of constructions that are part of the learned construction inventory is provided in Figure 4. The results show that holophrase constructions flourish in the earliest phase of the experiment. In a second phase, item-based and lexical constructions take over the role of the holophrase constructions, with an abundance of item-based constructions being created. Over the course of the experiment, the linguistic inventory of the learner gradually reaches a stable state consisting of a limited number of entrenched lexical constructions and (more general) item-based constructions. At the end of the experiment, the grammar consists on av-
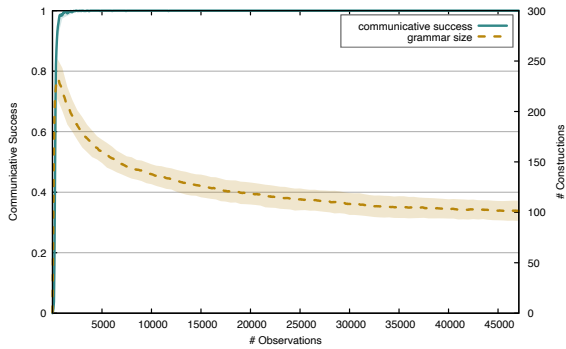
Figure 3: Evolution of communicative success (left y-axis) and grammar size (right y-axis) over time (full dataset).

erage of 10.2 holophrase constructions, 57.1 item-based constructions and 34.2 lexical constructions. These results show that the holophrase constructions have not yet completely disappeared after 47,134 observations and that the theoretical maximum of 35 lexical constructions was attained in 7 out of 10 experimental runs. Note that it is the dynamic evolution of the number of constructions per type over time that is important, rather than the absolute number of constructions at a given moment in time.
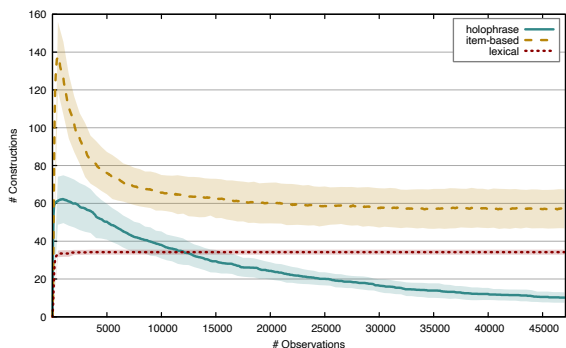

Figure 4: Evolution over time of the number of constructions per type with an entrenchment score > 0 (full dataset).

Figure 5 shows the active learning operators over time, zooming in on the first 1000 observations. In the beginning, only new holophrase constructions can be created. Then, operators of the first class can generalise over these holophrase constructions and create new item-based and lexical constructions. After that, operators of the second class take over and create constructions based on partial analyses. In the final phase of the experiment, mainly operators of the third class, which only create new categorial links, are active.
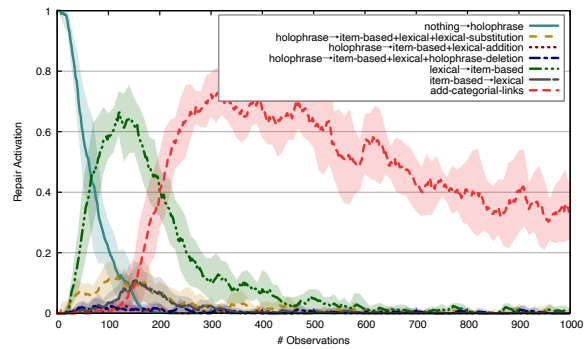

Figure 5: Active learning operators over time (first 1000 observations).

We finally conduct a secondary experiment, which consists in processing all observations from the test set using the grammars resulting from the different experimental runs of the primary experiment. The average communicative success amounts to a perfect 100% in both the comprehension and production direction. The average grammar size amounts to 101.5 constructions, of which 10.2 are holophrase constructions, 57.1 are item-based constructions and 34.2 are lexical constructions.

5 Related Work

Prior mechanistic models that operationalise the learning of constructions can be divided into two groups, based on the learning task that they address. A first class of models learns constructions paired with their meaning representation, either provided in the form of an annotated corpus (Dominey and Boucher, 2005; Chang, 2008; Abend et al., 2017) or obtained through task-oriented communicative interactions in a tutor-learner scenario (Gerasymova and Spranger, 2010; Beuls et al., 2010; Spranger and Steels, 2015). A second class of models, as introduced by Gaspers et al. (2011, 2016), is designed to learn form-meaning pairings under referential uncertainty. As such, the exact meaning representations of the input utterances are not provided to the learning algorithm, but grammars are learnt based on the combination of input utterances and situational context snippets. In these experiments, input utterances always correspond to a single term in the situational context. In general, both classes of models have explored interesting ideas on a rather small scale, either because they were limited to specific linguistic phenomena (Steels, 2004; Gerasymova and Spranger, 2010, 2012; Beuls et al., 2010; Spranger and Steels, 2015; Spranger, 2015, 2017;

Van Eecke and Beuls, 2017, 2018; Van Eecke, 2018), or because of the limited morpho-syntactic and semantic complexity of the input utterances (Dominey, 2005a,b, 2006; Chang, 2008; Gaspers et al., 2011; Gaspers and Cimiano, 2012, 2014; Gaspers et al., 2016; Abend et al., 2017). In all of the aforementioned work, either a segmentation of the input utterances, a lexicon or a set of predefined grammatical categories was provided. With the exception of the studies by Gaspers et al., the corpora that were used to learn and evaluate the models were not made available and were not described in sufficient detail to make reproduction and comparison feasible.

6 Discussion and Conclusion

The scientific contribution of the methodology and experiments presented in this paper is twofold. On the one hand, they provide computational evidence for the cognitive plausibility of constructivist theories of language acquisition. These theories, as most prominently put forward by Tomasello (2003), attribute the ability of children to acquire language to two main cognitive capacities: intention reading and pattern finding. Intention reading deals with reconstructing the intended meaning of observed utterances, while pattern finding implements generalisation processes that distil these reconstructed utterance-meaning pairs into abstract schemata embodying the linguistic knowledge of a language user. These schemata can then be used to fulfil the communicative function of language through the comprehension and production of natural language expressions. The methodology introduced in this paper presents a mechanistic model of the pattern finding capacity. Based on utterances paired with a representation of their meaning, the learning algorithm gradually builds up an inventory of concrete to abstract form-meaning mappings, called constructions, along with a network of emergent grammatical categories that captures how the constructions of the grammar can combine to collaboratively comprehend and produce utterances. The experiments show that a small number of general learning operators, which become active if an utterance cannot be successfully processed by the grammar learnt so far, effectively leads to learning dynamics that are similar to those described in the psycholinguistic literature (Pine and Lieven, 1997; Tomasello, 2003; Ambridge and Lieven, 2015). In the first phase of the learning

process, the learner acquires holistic mappings between utterances and their meaning representation. Soon after that, holophrase constructions are generalised to item-based constructions that integrate a variable slot. At the same time, this generalisation process leads to the emergence of slot-filling constructions, here called lexical constructions. Along with the item-based and lexical constructions, a network of grammatical categories emerges, capturing the distribution of construction slots and their observed fillers. In a third phase, more abstract item-based constructions emerge, with an increasingly large number of variable slots. In the final phase of the learning process, most constructions have already been acquired and most remaining impasses can be solved by adding new links to the categorial network. The learning dynamics are influenced by the degree of entrenchment of constructions. Constructions that are often successfully used become more entrenched, while their competitors are suppressed. As a result of this entrenchment process, the grammar reaches a stable state, while it remains adaptive to any changes in the discourse or environment. Similar dynamics have been observed in earlier experiments in the field of evolutionary linguistics, for instance in experiments on the emergence of compositionality in a population of autonomous agents (De Beule and Bergen, 2006; van Trijp, 2016).

On the other hand, the methodology and experiments presented in this paper pave the way for learning computationally tractable, large-scale, usage-based grammars that facilitate both language comprehension and production. The proposed learning algorithm supports online, interactive, incremental, transparent and data-efficient learning. The learner builds up its human-interpretable inventory of constructions and categories through the application of transparent syntactico-semantic generalisation processes. Already after a single observation, the fragment of linguistic knowledge acquired by the learner can be successfully used for language comprehension and production. As more and more utterance-meaning pairs are observed, the linguistic knowledge of the learner quickly expands and becomes better fit for achieving their communication goals. As a result of the dynamics of rewarding successful construction applications and punishing competing ones, the grammar of the learner remains ever-adaptive to any changes in the task or environment. Due to their online, interactive,

incremental, transparent and data-efficient nature, we strongly believe that the proposed mechanisms for learning computational construction grammars can serve as an excellent basis for implementing the language acquisition ability of truly intelligent agents.

Limitations

This paper has introduced a mechanistic model of the constructivist acquisition of language through syntactico-semantic pattern finding. Even though the results that we presented here proved to be promising and insightful, considerable challenges and limitations remain.

First of all, the learning operators that we present facilitate the learning of holistic, item-based and lexical constructions. At this point, the model does not include operators that give rise to constructions that capture more elaborate hierarchical patterns, including recursive patterns.

Second, the learning operators can adequately handle word order patterns, even non-contiguous ones. However, they provide no mechanisms to learn agreement patterns on an abstract level. As a consequence, different agreement patterns are captured in different constructions. This is a less-than-elegant solution, especially when applied to morphologically rich languages, as it can lead to a multiplication of the number of constructions.

Finally, the CLEVR dataset proved to be an excellent benchmark challenge for an initial validation of this novel methodology, as it consists of utterances with sufficient repetition, variation and overlap. It is however a synthetic dataset that does not reflect the richness of human language use. More research is needed before this methodology can be adequately applied to a broader range of linguistic resources, especially when it comes to finding generalisations over semantic structures.

Acknowledgements

References

Omri Abend, Tom Kwiatkowski, Nathaniel J. Smith, Sharon Goldwater, and Mark Steedman. 2017. Bootstrapping language acquisition. *Cognition*, 164:116–143.

Ben Ambridge and Elena Lieven. 2015. A constructivist account of child language acquisition. In Brian MacWhinney and William O'Grady, editors, *The Handbook of Language Emergence*, pages 478–510. John Wiley and Sons, Hoboken, NJ.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.

Katrien Beuls, Kateryna Gerasymova, and Remi van Trijp. 2010. Situated learning through the use of language games. In *Proceedings of the 19th Annual Machine Learning Conference of Belgium and The Netherlands (BeNeLearn)*, pages 1–6.

Katrien Beuls, Paul Van Eecke, and Vanja Sophie Cangalovic. 2021. A computational construction grammar approach to semantic frame extraction. *Linguistics Vanguard*, 7(1):20180015.

Jerome Bruner. 1983. *Learning to use language*. Oxford University Press, Oxford.

Nancy Chang. 2008. *Constructing grammar: A computational model of the emergence of early constructions*. Ph.D. thesis, University of California, Berkeley, Berkeley, CA.

Alonzo Church. 1932. A set of postulates for the foundation of logic. *Annals of Mathematics*, 33(2):346–366.

William Croft. 2000. *Explaining language change: An evolutionary approach*. Pearson Education, Harlow.

Joachim De Beule and Benjamin K. Bergen. 2006. On the emergence of compositionality. In *The Evolution of Language. Proceedings of the 6th International Conference (EVOLANG6)*, pages 35–42, Singapore. World Scientific.

Holger Diessel. 2004. *The acquisition of complex sentences*. Cambridge University Press, Cambridge.

Peter Ford Dominey. 2005a. Emergence of grammatical constructions: evidence from simulation and grounded agent experiments. *Connection Science*, 17(3-4):289–306.

Peter Ford Dominey. 2005b. From sensorimotor sequence to grammatical construction: Evidence from simulation and neurophysiology. *Adaptive Behavior*, 13(4):347–361.

Peter Ford Dominey. 2006. From holophrases to abstract grammatical constructions: Insights from simulation studies. In Eve Clark and Barbara Kelly, editors, *Constructions in acquisition*, pages 137–162. CSLI Publications, Stanford.

Peter Ford Dominey and Jean-David Boucher. 2005. Learning to talk about events from narrated video in a construction grammar framework. *Artificial Intelligence*, 167(1):31–61.

Judith Gaspers and Philipp Cimiano. 2012. A usage-based model for the online induction of constructions from phoneme sequences. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pages 1–6. IEEE.

Judith Gaspers, Philipp Cimiano, Sascha S Griffiths, and Britta Wrede. 2011. An unsupervised algorithm for the induction of constructions. In *2011 IEEE International Conference on Development and Learning (ICDL)*, volume 2, pages 1–6. IEEE.

Judith Gaspers, Philipp Cimiano, Katharina Rohlfing, and Britta Wrede. 2016. Constructing a language from scratch: Combining bottom–up and top–down learning processes in a computational model of language acquisition. *IEEE Transactions on Cognitive and Developmental Systems*, 9(2):183–196.

Judith Gaspers and Phlipp Cimiano. 2014. A computational model for the item-based induction of construction networks. *Cognitive Science*, 38 3:439–88.

Kateryna Gerasymova and Michael Spranger. 2010. Acquisition of grammar in autonomous artificial systems. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI-2010)*, pages 923–928.

Kateryna Gerasymova and Michael Spranger. 2012. An experiment in temporal language learning. In Luc Steels and Manfred Hild, editors, *Language Grounding in Robots*, pages 237–254. Springer, New York.

Adele Goldberg. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press, Chicago.

Adele Goldberg. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press, Oxford.

Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick. 2017. CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2901–2910.

Philip N Johnson-Laird. 1977. Procedural semantics. *Cognition*, 5(3):189–214.

Andrew Meltzoff. 1995. Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31(5):838–850.

Richard Montague. 1974. English as a formal language. In R. H. Thomason, editor, *Formal Philosophy; Selected papers of Richard Montague*, pages 188–221. Yale University Press, New Haven, CT.

Katherine Nelson. 1998. *Language in cognitive development: The emergence of the mediated mind*. Cambridge University Press, Cambridge.

Jens Nevens, Jonas Doumen, Paul Van Eecke, and Katrien Beuls. 2022. Language acquisition through intention reading and pattern finding. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 15–25, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Julian M Pine and Elena VM Lieven. 1997. Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18(2):123–138.

Dan Sperber and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Harvard University Press, Cambridge, MA.

Michael Spranger. 2015. Incremental grounded language learning in robot-robot interactions: Examples from spatial language. In *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pages 196–201.

Michael Spranger. 2017. Usage-based grounded construction learning: A computational model. In *The 2017 AAAI Spring Symposium Series*, pages 245–250, Palo Alto, Ca. AAAI Press.

Michael Spranger and Luc Steels. 2015. Co-acquisition of syntax and semantics: an investigation in spatial language. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 1909–1915, Palo Alto, Ca. AAAI Press.

Luc Steels. 2004. Constructivist development of grounded construction grammar. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pages 9–16.

Luc Steels, editor. 2011. *Design patterns in Fluid Construction Grammar*. John Benjamins, Amsterdam.

Michael Tomasello. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Harvard.

Michael Tomasello. 2009. The usage-based theory of language acquisition. In *The Cambridge handbook of child language*, pages 69–87. Cambridge University Press, Cambridge.

Paul Van Eecke. 2018. *Generalisation and specialisation operators for computational construction grammar and their application in evolutionary linguistics Research*. Ph.D. thesis, Vrije Universiteit Brussel, Brussels: VUB Press.

Paul Van Eecke and Katrien Beuls. 2017. Meta-layer problem solving for computational construction grammar. In *The 2017 AAAI Spring Symposium Series*, pages 258–265, Palo Alto, Ca. AAAI Press.

Paul Van Eecke and Katrien Beuls. 2018. Exploring the creative potential of computational construction grammar. *Zeitschrift für Anglistik und Amerikanistik*, 66(3):341–355.

Remi van Trijp. 2016. *The evolution of case grammar*. Language Science Press, Berlin.

Remi van Trijp, Katrien Beuls, and Paul Van Eecke. 2022. The FCG editor: An innovative environment for engineering computational construction grammars. *PLOS ONE*, 17(6):1–27.

Lara Verheyen, Jérôme Botoko Ekila, Jens Nevens, Paul Van Eecke, and Katrien Beuls. 2022. Hybrid procedural semantics for visual dialogue: An interactive web demonstration. In *Workshop on semantic techniques for narrative-based understanding: Workshop at IJCAI-ECAI 2022*, pages 48–52.

Tom Willaert, Paul Van Eecke, Katrien Beuls, and Luc Steels. 2020. Building social media observatories for monitoring online opinion dynamics. *Social Media + Society*, 6(2):2056305119898778.

Terry Winograd. 1972. Understanding natural language. *Cognitive Psychology*, 3(1):1–191.

A Appendix

A.1 Computing Infrastructure and Runtime

The experiments were run on a single 2.5 GHz CPU, with 16 GB RAM. Running a single series of the training set takes less than one hour. The processing of the test set takes less than 15 minutes.