

# The computational learning of construction grammars<sup>\*</sup>

## State of the art and prospective roadmap

Jonas Doumen<sup>1</sup>, Veronica Juliana Schmalz<sup>1</sup>, Katrien Beuls<sup>2</sup>  
and Paul Van Eecke<sup>1,3</sup>

<sup>1</sup> KU Leuven | <sup>2</sup> Université de Namur | <sup>3</sup> Vrije Universiteit Brussel

This paper documents and reviews the state of the art concerning computational models of construction grammar learning. It brings together prior work on the computational learning of form-meaning pairings, which has so far been studied in several distinct areas of research. The goal of this paper is threefold. First of all, it aims to synthesise the variety of methodologies that have been proposed to date and the results that have been obtained. Second, it aims to identify those parts of the challenge that have been successfully tackled and reveal those that require further research. Finally, it aims to provide a roadmap which can help to boost and streamline future research efforts on the computational learning of large-scale, usage-based construction grammars.

**Keywords:** construction grammar, computational construction grammar, usage-based linguistics, computational modelling, learning construction grammars

### 1. Learning computational construction grammars

The aim of this paper is to survey prior work on the computational learning of construction grammars, to identify gaps in the state of the art and to propose a perspective on the future of the field. The computational learning of construction grammars has traditionally been studied independently in different fields of research, in particular linguistics, cognitive science, computer science and artificial intelligence. As a consequence, research on this topic has been fragmented and interaction between the researchers involved has been scarce, as witnessed

---

\* Katrien Beuls and Paul Van Eecke are joint last authors.

by a lack of cross-referencing. In order to address this issue, this paper brings together the variety of methodologies that have been proposed in the literature and synthesises the results that have been achieved to date. Specifically, we define 14 criteria in light of which we review 31 models of construction grammar learning. We then identify important gaps in the state of the art and propose a roadmap that can help streamline future research efforts and investments. Our hope is to inspire a new generation of construction grammarians to boost progress on scaling usage-based constructionist approaches to language.

On a high level, computational models for learning construction grammars are motivated by three main reasons. A first motivation is theoretical in nature and concerns the computational operationalisation and validation of usage-based theories of language acquisition. As usage-based theories of language argue that languages are learnt through their use in communication, computational models of language acquisition can contribute crucial evidence in favour or against the scalability of specific theoretical arguments (see, e.g., Chang 2008). Second, the use of computational models brings important methodological advantages, as it enforces the use of precise and testable operational definitions, allows for the detection of theory-internal inconsistencies, and facilitates a fine-grained comparison of different theories. Such comparisons can play an important role in identifying inter-theoretical knowledge gaps and divergences (Bender 2008; Müller 2015). Finally, the machine learning of construction grammars is important from an application and valorisation perspective, as it facilitates the use of construction grammar insights and analyses in broad-domain language technology applications (van Trijp et al. 2022). Examples of such applications include visual question answering (see, e.g., Nevens et al. 2022; Verheyen et al. 2023), the frame-semantic analysis of discourse (see, e.g., Willaert et al. 2020; Beuls et al. 2021) and the construction-based analysis of corpora (see, e.g., EHAI 2023). Apart from direct applications, computational models of language acquisition are also relevant to the broader fields of computational linguistics and artificial intelligence as they provide a method for intelligent agents to learn to communicate through protocols that exhibit the robustness, flexibility and adaptivity of human languages (Van Eecke 2018; Beuls & Van Eecke 2023).

The remainder of this paper is structured as follows. Section 2 presents the criteria that were used to include prior research efforts into this literature review, as well as the criteria in light of which these efforts are described in Section 3. Section 4 synthesises the state of the art, identifies challenges and opportunities, and proposes a roadmap for future research. Section 5 concludes the paper.

## 2. Methodology

### 2.1 Inclusion criteria

The scope of this literature review concerns, to the best of our knowledge, all published prior research on the computational learning of construction grammars, as defined through the following inclusion criteria:

1. **Is the model concerned with learning form-meaning mappings?** This criterion concerns the nature of the linguistic knowledge that is learnt. Concretely, we only include models that concern the learning of form-meaning mappings. The selection is not influenced by the theoretical framework that is adopted nor by whether the authors situate their work within the field of construction grammar or not.
2. **Is the model computationally operationalised?** The second criterion concerns the operability of the learnt models. We only include models that have been implemented computationally.
3. **Can the model learn more than lexical mappings?** The third criterion concerns the nature of the constructions that are learnt. We only include models that acquire structures that transcend the level of individual words. For an overview of models that deal with vocabulary learning only, we refer the interested reader to Krenn et al. (2020).

As research within this scope has been carried in a variety of fields of research, theoretical frameworks, and scientific traditions, it was not feasible to design keyword searches or queries in scientific databases that would yield all relevant models. We came to the conclusion that any attempt to systematise the search process in such a manner left out important contributions that satisfy the inclusion criteria. Instead, we qualitatively scrutinised the literature to the best of our abilities and could include 31 models that satisfy the inclusion criteria.

### 2.2 Discussion criteria

Prior work on learning construction grammars stems from different fields of research and therefore adopts a wide variety of methods, terminologies, and experimental designs. In order to streamline the discussion and facilitate a meaningful comparison, we introduce 14 discussion criteria that will guide the review of the included models.

1. **Learning task.** Which learning task does the model address? Which problem is the model designed to solve? Which evaluation criteria are used?
2. **Dataset.** To which datasets has the model been applied?
3. **Input.** What is the nature of the input to the model?
4. **Form complexity.** What is the morpho-syntactic complexity of the linguistic input?
5. **Meaning complexity.** What is the semantic complexity of the linguistic input?
6. **Grounding.** Is the meaning representation grounded in a situation model?
7. **Segmentation level.** What level of segmentation of the input is provided to the model (phonemes, graphemes, words, or utterances)?
8. **Lexicon.** Is a predefined lexicon provided?
9. **Grammatical categories.** Is a predefined set of grammatical categories provided?
10. **Incremental learning.** Does the model learn in an incremental fashion, i.e., dynamically extending its knowledge after each exemplar?
11. **Bi-directional grammar.** Can the learnt grammar be used for both language comprehension and production?
12. **Abstraction level.** What is the level of abstraction of the learnt constructions? Is the grammar limited to item-based constructions and slot-filling constructions or does it include hierarchical and recursive constructions?
13. **Non-compositionality.** Is the learning algorithm able to capture non-compositional aspects of language use? Is all meaning lexicalised or can grammatical constructions also contribute semantic information?
14. **Benchmark.** Is the data that was used during the learning and evaluation process precisely described and available to the community?

Apart from guiding the literature review, the discussion criteria are also used to organise the schematic synthesis of the literature presented in Table 1. Here, qualitative descriptions are provided for *learning task*, *dataset*, *input*, *form complexity* and *meaning complexity*. Boolean values are used to assess the criteria of *grounding*, *lexicon*, *grammatical categories*, *incremental learning*, *bi-directional grammar*, *non-compositionality* and *benchmark*. *Segmentation level* can either be ‘phonemes’, ‘graphemes’, or ‘words’, and *abstraction level* can either be ‘item-based’, or ‘hierarchical’, with only the highest level of abstraction being listed. The information contained in the table reflects the properties of the models as presented in the papers that introduce them, without implying that any limitations are inherent to the methodologies.

### 3. Review of prior literature

This section provides a detailed review of all included models. A comparative analysis of all models in terms of the discussion criteria introduced in Section 2.2 is also provided in Table 1. Readers who are primarily interested in the higher-level picture can safely skip ahead to Section 4, where we present a synthesis of the details covered in this section.

On the highest level, the models can be organised according to the overall learning task that they are designed to tackle:

1. **Learning a maximally concise grammar.** The task is to find a minimal set of constructions that optimally covers a corpus of language use.
2. **Learning a grammar from utterance-meaning pairs.** The task is to learn a grammar that maps between utterances and their meaning representation, whereby a gold semantic annotation is provided for each utterance.
3. **Learning a grammar under referential uncertainty.** The task is to learn a grammar that maps between utterances and their meaning representation, whereby a superset of the gold semantic annotation is provided. The referential uncertainty stems from the fact that the exact meaning representation is not provided.
4. **Learning a grammar from a situation model.** The task is to learn a grammar that maps between utterances and their meaning representation, whereby no gold semantic annotation is provided. The meaning has to be abductively derived from a situation model.

The following sections reflect the organisation of the models in terms of these four learning tasks.

Table 1. Comparative analysis of the literature in terms of the discussion criteria proposed in Section 2.2

| Learning Task               | Dataset                                          | Input                                                                     | Form complexity                                   |                                                               | Semantic complexity                                  |   | Grounded |   | Categories given | Segmentation level | Incremental | Bi-directional | Abstraction level | Non-compositionality | Benchmark |
|-----------------------------|--------------------------------------------------|---------------------------------------------------------------------------|---------------------------------------------------|---------------------------------------------------------------|------------------------------------------------------|---|----------|---|------------------|--------------------|-------------|----------------|-------------------|----------------------|-----------|
|                             |                                                  |                                                                           |                                                   |                                                               |                                                      |   |          |   |                  |                    |             |                |                   |                      |           |
| Abend et al. (2017)         | learning a grammar under referential uncertainty | CHILDES Eve section, with automatically generated meaning representations | utterances with candidate meaning representations | simple utterances (intransitives and transitives)             | one predicate per utterance                          | N | N        | Y | Y                | words              | Y           | N              | hierarchy         | N                    | N         |
| Alishahi & Stevenson (2008) | learning a grammar from utterance-meaning pairs  | English synthetic data                                                    | utterance-meaning pairs                           | simple utterances (intransitives, transitives, ditransitives) | one predicate per utterance                          | N | Y        | Y | Y                | words              | Y           | Y              | item-based        | N                    | N         |
| Artzi & Zettlemoyer (2013)  | learning a grammar from a situation model        | Navi corpus                                                               | utterance, start state and validation function    | navigation instructions with spatial relations                | lambda calculus with Neo-Davidsonian event semantics | Y | Y        | Y | Y                | words              | N           | N              | hierarchy         | N                    | Y         |
| Beekhuizen & Bod (2014)     | learning a grammar under referential uncertainty | English synthetic data                                                    | utterances with candidate meaning representations | simple utterances (intransitives and transitives)             | one predicate per utterance                          | N | N        | Y | Y                | words              | Y           | N              | hierarchy         | Y                    | N         |

Table 1. (continued)

| Learning Task          |                                                  | Dataset                                                               | Input                                             | Form complexity                                       |                             | Semantic complexity         |   |   | Grounded | Lexicon given | Categories given | Segmentation level | Incremental | Bi-directional | Abstraction level | Non-compositionality | Benchmark |
|------------------------|--------------------------------------------------|-----------------------------------------------------------------------|---------------------------------------------------|-------------------------------------------------------|-----------------------------|-----------------------------|---|---|----------|---------------|------------------|--------------------|-------------|----------------|-------------------|----------------------|-----------|
| Beekhuizen (2015)      | learning a grammar under referential uncertainty | English synthetic data                                                | utterances with candidate meaning representations | simple utterances (intransitives and transitives)     | one predicate per utterance | one predicate per utterance | N | Y | N        | Y             | words            | Y                  | Y           | Y              | hierarchy         | Y                    | N         |
|                        | learning a grammar from utterance-meaning pairs  |                                                                       |                                                   |                                                       |                             |                             |   |   |          |               |                  |                    |             |                |                   |                      |           |
| Beuls et al. (2010)    | learning a grammar from utterance-meaning pairs  | Hungarian synthetic data                                              | utterance-meaning pairs                           | simple utterances with poly-personal verbal agreement | one predicate per utterance | one predicate per utterance | N | Y | Y        | words         | Y                | Y                  | Y           | item-based     | Y                 | N                    | N         |
|                        | learning a grammar from utterance-meaning pairs  |                                                                       |                                                   |                                                       |                             |                             |   |   |          |               |                  |                    |             |                |                   |                      |           |
| Chang (2008)           | learning a grammar from utterance-meaning pairs  | CHILDES Naomi section with manually annotated meaning representations | utterance-meaning pairs                           | simple utterances (intransitives and transitives)     | one predicate per utterance | one predicate per utterance | N | Y | Y        | words         | Y                | Y                  | N           | item-based     | Y                 | N                    | N         |
|                        | learning a grammar under referential uncertainty |                                                                       |                                                   |                                                       |                             |                             |   |   |          |               |                  |                    |             |                |                   |                      |           |
| Chen and Mooney (2008) | learning a grammar under referential uncertainty | RoboCup                                                               | utterances with candidate meaning representations | simple utterances (intransitives and transitives)     | one predicate per utterance | one predicate per utterance | Y | N | N        | N             | words            | N                  | Y           | unk.           | N                 | Y                    | Y         |

Table 1. (continued)

| Learning Task            | Dataset                                         | Input                       | Form complexity                                                   | Semantic complexity                                           | Grounded                          | Lexicon given | Categories given | Segmentation level | Incremental | Bi-directional | Abstraction level | Non-compositionality | Benchmark |
|--------------------------|-------------------------------------------------|-----------------------------|-------------------------------------------------------------------|---------------------------------------------------------------|-----------------------------------|---------------|------------------|--------------------|-------------|----------------|-------------------|----------------------|-----------|
| Dominey (2005a)          | learning a grammar from utterance-meaning pairs | English synthetic data      | utterance-meaning pairs (derived from object and event detection) | simple utterances (intransitives, transitives, ditransitives) | max. two predicates per utterance | Y             | Y                | Y                  | words       | Y              | N                 | item-based           | N         |
| Dominey & Boucher (2005) | learning a grammar from utterance-meaning pairs | English synthetic data      | utterance-meaning pairs (derived from object and event detection) | simple utterances (intransitives, transitives, ditransitives) | one predicate per utterance       | Y             | N                | Y                  | words       | N              | Y                 | item-based           | N         |
| Doumen et al. (2023)     | learning a grammar from utterance-meaning pairs | CLEVR                       | utterance-meaning pairs                                           | simple queries about properties of objects                    | procedural semantic networks      | N             | N                | N                  | words       | Y              | Y                 | item-based           | Y         |
| Dunn (2017)              | learning a maximally concise grammar            | English web data from ukWac | POS and semantic tags                                             | full (broad coverage)                                         | semantic tags                     | N             | Y                | Y                  | words       | N              | N                 | hierarchy            | Y         |





Table 1. (continued)

| Learning Task            |                                                  | Dataset | Input                                             | Form complexity                                   | Semantic complexity         | Grounded | Lexicon given | Categories given | Segmentation level | Incremental | Bi-directional | Abstraction level | Benchmark |
|--------------------------|--------------------------------------------------|---------|---------------------------------------------------|---------------------------------------------------|-----------------------------|----------|---------------|------------------|--------------------|-------------|----------------|-------------------|-----------|
| Gaspers et al. (2011)    | learning a grammar under referential uncertainty | RoboCup | utterances with candidate meaning representations | simple utterances (intransitives and transitives) | one predicate per utterance | N        | N             | N                | words              | N           | N              | item-based        | unk. Y    |
| Gaspers & Cimiano (2012) | learning a grammar under referential uncertainty | RoboCup | utterances with candidate meaning representations | simple utterances (intransitives and transitives) | one predicate per utterance | N        | N             | N                | phon.              | Y           | N              | item-based        | unk. Y    |
| Gaspers & Cimiano (2014) | learning a grammar under referential uncertainty | RoboCup | utterances with candidate meaning representations | simple utterances (intransitives and transitives) | one predicate per utterance | N        | N             | N                | phon.              | Y           | Y              | item-based        | unk. Y    |
| Gaspers et al. (2016)    | learning a grammar under referential uncertainty | RoboCup | utterances with candidate meaning representations | simple utterances (intransitives and transitives) | one predicate per utterance | N        | N             | N                | phon.              | Y           | N              | item-based        | unk. Y    |



Table 1. (continued)

| Learning Task             | Dataset                                          | Input                                                                     | Form complexity                                   |                                                   | Semantic complexity          | Grounded                    |               |                  | Segmentation level | Incremental | Bi-directional | Abstraction level | Non-compositionality | Benchmark |
|---------------------------|--------------------------------------------------|---------------------------------------------------------------------------|---------------------------------------------------|---------------------------------------------------|------------------------------|-----------------------------|---------------|------------------|--------------------|-------------|----------------|-------------------|----------------------|-----------|
|                           |                                                  |                                                                           | utterances with candidate meaning representations | simple utterances (intransitives and transitives) |                              | one predicate per utterance | Lexicon given | Categories given |                    |             |                |                   |                      |           |
| Kwiatkowski et al. (2012) | learning a grammar under referential uncertainty | CHILDES Eve section, with automatically generated meaning representations | utterances with candidate meaning representations | simple utterances (intransitives and transitives) | one predicate per utterance  | N                           | N             | Y                | words              | Y           | N              | hierarchy         | N                    | N         |
| Martí et al. (2021)       | learning a maximally concise grammar             | Spanish web data                                                          | POS and dependency labels                         | full (broad coverage)                             | semantic tags                | N                           | Y             | Y                | words              | N           | N              | hierarchy         | Y                    | Y         |
| Nevens et al. (2022)      | learning a grammar from a situation model        | CLEVR                                                                     | questions and feedback                            | simple queries about properties of objects        | procedural semantic networks | Y                           | N             | N                | words              | Y           | Y              | item-based        | Y                    | Y         |
|                           |                                                  |                                                                           |                                                   |                                                   |                              |                             |               |                  |                    |             |                |                   |                      |           |
| Spranger & Steels (2015)  | learning a grammar from a situation model        | English synthetic data                                                    | questions and feedback                            | simple utterances expressing spatial relations    | procedural semantic networks | Y                           | Y             | Y                | words              | Y           | Y              | item-based        | Y                    | N         |

Table 1. (continued)

|                 | Learning Task                                   | Dataset                | Input                                                             | Form complexity                                |                              | Semantic complexity |   | Grounded Lexicon given | Categories given | Segmentation level | Incremental | Bi-directional | Abstraction level | Non-compositionality | Benchmark |
|-----------------|-------------------------------------------------|------------------------|-------------------------------------------------------------------|------------------------------------------------|------------------------------|---------------------|---|------------------------|------------------|--------------------|-------------|----------------|-------------------|----------------------|-----------|
|                 |                                                 |                        |                                                                   |                                                |                              |                     |   |                        |                  |                    |             |                |                   |                      |           |
| Spranger (2015) | learning a grammar from a situation model       | English synthetic data | questions and feedback                                            | simple utterances expressing spatial relations | procedural semantic networks | Y                   | Y | Y                      | Y                | words              | Y           | Y              | item-based        | Y                    | N         |
| Spranger (2017) | learning a grammar from a situation model       | English synthetic data | questions and feedback                                            | simple utterances expressing spatial relations | procedural semantic networks | Y                   | Y | Y                      | Y                | words              | Y           | Y              | item-based        | Y                    | N         |
| Steels (2004)   | learning a grammar from utterance-meaning pairs | English synthetic data | utterance-meaning pairs (derived from object and event detection) | simple utterances                              | predicate calculus           | Y                   | N | N                      | N                | words              | Y           | Y              | hierarchy         | Y                    | N         |

### 3.1 Learning a maximally concise grammar

The first category of models addresses the task of finding a minimal set of constructions that optimally covers a corpus of language use. Dunn (2017) introduces a method to induce schematic patterns from large amounts of web-crawled corpus data. These patterns take the form of a sequence of slots that can be filled by word forms, morpho-syntactic categories or semantic categories. These categories are provided in the form of annotation layers. The resulting grammars are evaluated against a held-out test set in terms of various measures, including minimal description length and  $\Delta P$  entrenchment (Ellis 2006). Variations on the method are described by Dunn (2018, 2019, 2022, 2023) and Dunn & Tayyar Madabushi (2021). Martí et al. (2021) present DISCO, a methodology for discovering constructional candidates in large web-crawled corpora of Spanish texts. Similar to Dunn (2017), the semantic categories integrated into the ‘lexico-syntactic’ patterns are modelled as clusters over the distributional semantic representations of lemmata. The retrieved patterns are evaluated using statistical association measures as well as through manual evaluation by expert linguists.

This line of work models the induction of partially abstract patterns that combine form-related and meaning-related features. While relevant from a construction grammar perspective, the resulting patterns do not correspond to constructions that actually constitute mappings between aspects of form and meaning. These models thereby do not support mapping between utterances and their meaning representations.

### 3.2 Learning a grammar from utterance-meaning pairs

The second category of models addresses the task of learning a construction grammar from utterance-meaning pairs.

Dominey (2005a, 2005b, 2006) and Dominey & Boucher (2005) present a neural model for the acquisition of holophrase constructions, item-based constructions and abstract constructions that capture argument structure relations (e.g., transitives and ditransitives). Learners start with the capability to distinguish between closed-class and open-class words and learn to map between slots in the argument structure constructions and the semantic roles they take. Learning the meaning of open-class words is tackled as an initial cross-situational learning step. Item-based constructions are then learned from the remaining closed-class words (e.g., *X WAS Y TO Z BY A*) by storing the mapping of the order of thematic roles (e.g., object, action, recipient, agent) and these closed-class elements as constructions in the construction inventory.

Alishahi & Stevenson (2008) present a computational model that mimics the acquisition of verb argument structure. The input utterance-meaning pairs are generated based on the 20 most frequent verbs of a subsection of the CHILDES corpus (MacWhinney 2000). An example would be the utterance *Mom put toys in boxes* paired to the meaning representation  $\text{PUT}_{[\text{CAUSE}, \text{MOVE}]}(\text{MOM} \langle \text{AGENT} \rangle, \text{TOYS} \langle \text{THEME} \rangle, \text{IN}_{[]}(\text{BOXES} \langle \text{DESTINATION} \rangle) \langle \text{DESTINATION} \rangle)$ . The lexicon and semantic roles are given to the model a priori. A construction is defined as a probabilistic association between syntactic and semantic features of a verb and its arguments, which emerged over a number of observations. Language processing is tackled as a Bayesian prediction problem. In production, the model predicts the syntax from a given semantic structure, and in comprehension, it predicts a (partial) semantic structure for a given utterance. The authors present experiments for both comprehension and production. The experiments show that the model exhibits effects of syntactic and semantic bootstrapping, that it can recover from over-generalisation, and that it is robust to noise, thereby mimicking the stages of child language acquisition.

Chang (2008) presents a set of learning operators that operationalise Bayesian model merging in the framework of Embodied Construction Grammar (ECG). The task is to learn a grammar that can correctly comprehend utterances from a held-out test set at certain intervals during training. The dataset that is used consists in a schema-annotated subset of the CHILDES corpus, containing parent-child interactions from a child from 15 until 24 months of age. A lexicon and ontology are provided, corresponding to the linguistic knowledge of a child at the two-word stage. Two classes of learning operators are introduced: mapping operators and reorganisation operators. In the first class of operators, the simple mapping operator maps either all, or a part of the uncovered form to its co-occurring context information. Relational mapping creates a new construction that encodes the syntactic and semantic relations between existing constructions that could apply. The applied constructions become constituents in the newly formed item-based construction. For the input utterance *throw ball* for example, the lexical constructions for *THROW* and *BALL* may already be known. A new construction is then created that encodes that *THROW* precedes *BALL*, and that the ball is the throwee in the throwing event. Next, three operators can apply: merging, joining and splitting. The merge operator learns a generalised construction for two constructions that contain shared structure. For example, *THROW-BALL* and *THROW-BLOCK* can be generalised to *THROW-TOY*. A second reorganisation operator joins existing constructions that have a certain overlap, for example *HUMAN-THROW* and *THROW-BOTTLE* can be joined into *HUMAN-THROW-BOTTLE*. Inversely, the splitting operator creates new constructions by taking the set difference between existing constructions, for example *THROW-FRISBEE* and *THROW*

can be split into FRISBEE, and a new item-based construction, THROW-*x*, which takes the newly created FRISBEE-construction as a constituent. After the application of each applicable learning operator for a given input, the model uses minimum description length to select the shortest grammar that covers the data. The model is evaluated in relation to a hand-crafted gold standard grammar in terms of form and meaning coverage, and description length. The presented experiments operate at a rather small scale: first on 40 tokens of the verbs *fall* and *throw*, then on 200 tokens of caused-motion and self-motion predicates. The author reports that, in qualitative evaluation, both experiments show a resemblance to the stages described in the literature of usage-based language acquisition, i.e., going from concrete to increasingly abstract constructions.

Kwiatkowski et al. (2010) introduce a unification-based learning algorithm, which is designed to induce semantic parsers from corpora consisting of sentence and logical-form pairs. The model learns a lexicon and parameters in Combinatory Categorical Grammar (CCG). In CCG, a fixed set of combinatory rules is given, which govern how both categories and arguments of logical predicates can be combined. A lexical entry consists of a word form, its syntactic category and a logical predicate, for example: *what*  $\vdash S / NP : \lambda x.answer(x)$ . The model initially creates a holistic construction by mapping an entire sentence to its logical representation. The main generalisation operator splits these overly specific entries into smaller, more widely applicable lexical entries. Higher-order unification is used to restrict the number of hypotheses, while safeguarding the integrity and combinatory properties of the logical forms. As there are many possible analyses for a given sentence, a log-linear model is used to select the most likely analysis. Before training, the model is initialised with a list of NPs that are present in the data (e.g., *Texas*  $\vdash NP : tex$ ). The model is evaluated for four languages and two semantic representation formats on the GeoQuery dataset (Zelle & Mooney 1996). Kwiatkowski et al. (2011) apply a further iteration of the same methodology to the more natural Air Travel Information System pilot corpus (ATIS – Hemphill et al. 1990).

In contrast to the other work presented in this section, CCG assumes that meaning is fully lexicalised, and thereby that language is fully compositional.

Gerasymova & Spranger (2010, 2012) investigate the acquisition of holophrase constructions, item-based constructions and abstract constructions, represented and processed using the Fluid Construction Grammar (FCG) framework (Steels & De Beule 2006; van Trijp et al. 2022; Beuls & Van Eecke 2023), for Russian aspectual marking in a tutor-learner language game setting (Steels 1998, 2001). Holophrase constructions are learnt by a straightforward mapping operation between an observed form and its meaning. Item-based constructions and abstract constructions are learnt as generalisations over pre-categorised lexical



items. Beuls et al. (2010) apply the same methodology to the conjugation of verbs in Hungarian, with a special focus on its intricate agreement marking system. Garcia Casademont & Steels (2015, 2016) and Garcia Casademont (2018) apply the same paradigm to the acquisition of abstract, hierarchical and recursive constructions.

Doumen et al. (2023) present a model of the co-acquisition of grammatical categories and form-meaning mappings, ranging from lexical to item-based constructions. The model operationalises Tomasello's (2003) pattern finding mechanisms as a set of meta-layer learning operators in Fluid Construction Grammar. The base operator stores an entire utterance with its meaning representation as a holophrase construction. A second class of operators generalises over holophrase constructions and new observations. When a difference in both form and meaning is found, these operators learn the syntactic and semantic composition of the observation by means of substitution, addition and deletion operations, resulting in item-based constructions, lexical constructions and a network of emergent grammatical categories that governs how constructions can combine. A third class of operators starts from partial analyses. When an observation is partially covered by one or multiple constructions in the agent's inventory, the model applies these to the utterance, and a new item-based or lexical construction is learned, along with corresponding categorial links. A final learning operator adds new categorial links if all constructions that are necessary to cover an utterance are known, but no links in the categorial network are found. The model is evaluated on a subset of the CLEVR dataset (Johnson et al. 2017).

### 3.3 Learning a grammar under referential uncertainty

The third category of models addresses the task of learning a construction grammar under referential uncertainty. While the previous category of models took as input utterances paired with their semantic representation, this category of models takes as input utterances paired with a superset of their semantic representation. The fact that the meaning representation is noisy thereby poses an additional challenge to the learning task.

A first line of work in this category focuses on aligning commentaries with observed actions in Robocup football games, a task introduced by Chen & Mooney (2008). The input data concerns a list of natural language utterances of limited morpho-syntactic complexity, each paired with a number of candidate meaning representations. The correct meaning representation for an utterance always consists in a single predicate. The task is then to learn a productive grammar that can map between utterances and their meaning representation. Along with the task and corpus, Chen & Mooney (2008) also present a method to induce

*probabilistic synchronous context-free grammars* that can effectively be used for semantic parsing and meaning-driven language production. Gaspers et al. (2011) make use of the same task and corpus to develop and evaluate a method to learn construction grammars without strict supervision. In a first phase, a probabilistic lexicon is learned using a cross-situational learning algorithm (Fazly et al. 2010). The lexicon is restricted to sequences of one or two tokens, paired with one predicate or argument that occurs in the corpus. In a second phase, schemata are computed as generalisations over co-occurring lexical items and predicates or arguments. In particular, lexical items appearing in the same syntactic environments are grouped into *semantic equivalence classes*. Then, schemata are constructed by substituting the lexical items with abstract labels associated to their respective semantic equivalence class. Gaspers & Cimiano (2012, 2014) and Gaspers et al. (2016) present variations on this model, where the level of segmentation of the input is reduced from tokens through graphemes to phonemes.

Introducing the framework of Meaningful Unsupervised Data Oriented Parsing ( $\mu$ -DOP), Beekhuizen & Bod (2014) present a model that incrementally learns mappings between nodes in binary trees and meaning representations expressed in predicate logic. The task that the model aims to solve is to find the correct predicate-argument structure for a given utterance in referentially ambiguous scenes. They apply the model to artificial data that is not fully compositional on the meaning side, including for example idioms. At the start of an experiment, the construction inventory is empty. When presented with a new utterance, the model first attempts to apply any existing constructions. In order to induce syntactic structures, the model first generates all possible binary trees for a given utterance. Similarly, on the meaning side, the model decomposes the meaning predicates into all possible decompositions. The arguments are replaced by slot indices. For example  $\text{WATCH}(E_1, E_2)$  can be further decomposed into  $P(E_1, E_2)$  and  $P$ :  $\text{WATCH}$ . All unanalysed nodes in the binary trees are then paired with all possible decompositions of parts of the semantic representation as candidate form-meaning pairs. These are then combined with the possible binary tree analyses to form a hypothesis space of possible analyses. As a last step, the model proceeds to extract all subtrees, and compares them to previously observed subtrees. If a subtree is found in multiple utterances' parse trees, then its prior probability of being a valid constituent increases. Non-compositional parts are retained by only allowing further decomposition at nodes where a meaning representation is present. In both parsing and production, the derivation with the highest joint prior probability is selected, which in practice means that hypotheses with a smaller amount of subtrees are preferred. As a further evolution of this work, Beekhuizen et al. (2014) and Beekhuizen (2015) introduce the Syntagmatic-Paradigmatic Learner (SPL), which addresses a number of desiderata inspired by

usage-based theories of language acquisition. As such, the generation of possible analyses, called derivations, is driven by a set of general structure-building operations, implementing parts-to-whole learning strategies. These operations allow for the concatenation of derivations, slot filling, syntactic bootstrapping, and ignoring words. The model is now applied to artificial data that is generated based on empirical research on noise, uncertainty, and situational continuation in child-directed communicative interactions.

Kwiatkowski et al. (2012) and Abend et al. (2017) introduce a probabilistic learning algorithm for combinatory categorial grammars (CCG), designed to model the human language acquisition process. This approach differs from earlier work by Kwiatkowski et al. (2010), presented in Section 3.2, in a few key aspects. An evident difference is that this model learns from utterance-situation pairs under propositional uncertainty, rather than having a single gold standard semantic representation for each utterance. A second difference is that the model is able to learn the lexicon, and the mapping between the lexicon and predefined categories from scratch, rather than relying on a list of noun phrases. An important assumption in this line of work is that a language-independent set of combinatory rules is provided to the model. The grammar is evaluated both quantitatively and qualitatively using the Eve corpus (Brown 1973), which was enriched with deterministically mapped logical forms based on dependency tree annotations.

### 3.4 Learning a grammar from a situation model

The final category of models addresses the task of learning a construction grammar from utterances observed during communicative interactions. As such, the meaning representation of the utterances is not provided, but needs to be abductively inferred from the situation in which the interaction takes place.

Steels (2004) presents an initial experiment in which a population of artificial agents bootstrap argument structure constructions and grammatical categories from visually grounded meaning representations. The paper presents a description game, in which two agents from the population observe a scene that needs to be successfully described by one agent to the other. The scene is transcribed in terms of first order logic predicates. During the conceptualisation phase, the speaker agent decides on the ‘event profile’ that they want to express (see, e.g., Croft 1998), i.e., deciding which roles have to be expressed linguistically. Whether speaking or listening, an agent first makes use of its own construction grammar to process the conceptualised meaning representation or observed utterance. When an agent fails to formulate an utterance that expresses the conceptualised meaning representation or fails to comprehend an observed utterance in terms of the current scene, a game fails and a learning event takes place. When a game succeeds, the score of an

agent's applied constructions is increased, while the score of competing constructions is decreased. Upon failure, the scores of the constructions that were used are decreased. Overall, the presented methodology shows how hierarchical semantic and syntactic categories can be learnt, and how the emergence of syntax aids to resolve ambiguity. van Trijp (2008, 2016) presents an extensive suite of follow-up experiments and shows how abstract semantic roles can emerge and evolve in populations of autonomous agents through multi-level selection strategies.

Artzi & Zettlemoyer (2013) present a model where a seed lexicon, a predefined set of combinatory rules and a situation model are provided. The seed lexicon and combinatory rules are used to generate hypotheses about the meaning underlying observed utterances, which can subsequently be validated against the situation model. The goal is to extend the seed lexicon with new items in order to solve a navigation task.

Spranger & Steels (2015) and Spranger (2015, 2017) present a model of the acquisition of spatial language in embodied artificial agents. In a shared environment and through a tutor-learner language game, two agents interact with 15 different objects in over 1000 different spatial scenes. The goal is to simultaneously acquire the semantic and syntactic aspects of spatial language. Each communicative interaction proceeds as follows. The tutor agent formulates an utterance that uniquely refers to an object in the situational context. The learner agent comprehends and interprets the utterance with respect to the scene and points to the object that results from the interpretation process. Then, feedback is provided by the tutor in the form of pointing. If the learner agent misinterpreted the observed utterance, it needs to make a hypothesis about the intended meaning of the utterance. The agent does this by composing a procedural semantic network based on a set of cognitive operations that it can perform. The observed form and hypothesised meaning can then be stored in the form of a holophrastic construction. Later, the agent can generalise over the constructions it knows and thereby construct more abstract constructions. The semantic classes of the slots in the more abstract constructions are predefined in an ontology that the learner can access.

Nevens et al. (2022) build further on the composition processes introduced by Spranger & Steels (2015) and the syntactico-semantic generalisation operators introduced by Doumen et al. (2023). They tackle the visual question answering benchmark introduced by Johnson et al. (2017) and are able to bootstrap a construction grammar that maps between English questions and visually grounded queries without ever having observed the queries. Through lateral inhibition alignment dynamics, generally applicable constructions become gradually more entrenched, while suboptimal constructions, i.e., constructions that do not generalise to other scenes, for example due to suboptimal hypotheses about the intended meaning of an observed utterance, gradually disappear from the grammar.

## 4. Discussion

The low-level review of the prior literature in the previous section reveals perhaps most clearly that existing models for computationally learning construction grammars are highly diverse in nature and therefore challenging to compare. Their diversity is situated on almost any level, from the task that is tackled, to the goals that are envisioned, the datasets that are used, the approaches that are taken and the methodologies that are applied. So, where are we at now and where should we be heading?

Answering these questions presupposes that a particular perspective is taken. After all, researchers in natural language processing, for example, have different immediate goals and concerns than researchers in cognitive science or language pedagogy. We will be addressing these questions from a constructionist perspective, with the general goal in mind of operationalising large-scale, usage-based construction grammar. So, what does it mean to operationalise usage-based construction grammar on a large scale? We define the term *operational* as having a computational implementation that supports the processes of language comprehension (i.e., mapping from an utterance to a representation of its meaning) and production (i.e., mapping from a meaning representation to an utterance that expresses it). The term *large-scale* can be interpreted as having a broad and domain-general coverage of the language. For the term *usage-based*, we adhere to Bybee's (2006) view that an individual's grammar is rooted in and shaped by the individual's history of communicative interactions. Finally, we interpret the term *construction grammar* as adhering to the basic principles underlying constructionist approaches to language (Goldberg 2003). These principles are summarised by van Trijp et al. (2022) as follows: (i) all linguistic knowledge is captured in the form of constructions, i.e., form-meaning pairings, (ii) there is no strict distinction between words and grammar rules, so that non-compositional aspects of the language can elegantly be captured, (iii) constructions cut across the different levels of linguistic analysis, and (iv) construction grammars are learnt, dynamic systems.

An important insight gained through the literature review is that prior models typically score well on one or two of these dimensions, while suffering from substantial shortcomings on the others. As such, the models that address the *large-scale* dimension of the challenge, i.e., those by Dunn (2017, 2018, 2019, 2022, 2023), Dunn & Tayyar Madabushi (2021) and Martí et al. (2021), do not yield individual grammars that support language comprehension and production. The scale of other models is restricted in one of three ways. Some models cover specific linguistic phenomena only, such as basic argument structure (Steels 2004; Dominey 2005a, 2005b, 2006; Dominey & Boucher 2005; van Trijp 2008, 2016; Garcia

Casademont & Steels 2015, 2016; Garcia Casademont 2018), aspectual marking (Gerasymova & Spranger 2010, 2012), agreement marking (Beuls et al. 2010) or spatial relations (Spranger & Steels 2015; Spranger 2015, 2017). Other models are applied to narrow artificial datasets, such as Robocup (Chen & Mooney 2008; Gaspers et al. 2011; Gaspers & Cimiano 2012, 2014; Gaspers et al. 2016), Geo-Query (Kwiatkowski et al. 2010, 2011), CLEVR (Nevens et al. 2022; Doumen et al. 2023), ATIS (Kwiatkowski et al. 2011), Navi (Artzi & Zettlemoyer 2013), or other datasets specifically generated for evaluating the model (Beekhuizen & Bod 2014; Beekhuizen et al. 2014; Beekhuizen 2015). A final set of models makes use of small-scale corpora of transcribed children's speech (Alishahi & Stevenson 2008; Chang 2008; Kwiatkowski et al. 2012; Abend et al. 2017). The large-scale operationalisation of construction grammar learning thereby remains an open challenge.

When it comes to the *usage-based* nature of the models, we can define a broad spectrum of possible approaches. On the most usage-based side of the spectrum, we would expect to find models that learn individual construction grammars from empirically observed communicative interactions, without access to a segmentation of the utterances, their meaning representations, or any predefined categories or lexical items. No models that satisfy all of these criteria were found in the literature. In fact, all prior models learn from utterances that are segmented on some level (phonemes, graphemes or words). Existing models that learn individual grammars without any access to annotated meaning representations are always learnt based on artificial utterances (Nevens et al. 2022), sometimes in combination with a predefined lexicon and system of categories (Artzi & Zettlemoyer 2013; Spranger & Steels 2015; Spranger 2015, 2017). Other models that do not have access to exact meaning representations do have access to a superset of the meaning representations. These models learn from artificial utterances (Beekhuizen & Bod 2014; Beekhuizen et al. 2014; Beekhuizen 2015) or short natural utterances (Chen & Mooney 2008; Gaspers et al. 2011; Gaspers & Cimiano 2012, 2014; Gaspers et al. 2016), sometimes in combination with a set of predefined combinatory rules (Kwiatkowski et al. 2012; Abend et al. 2017). Next, we find a class of models that learn from semantically annotated utterances. Some of these models learn construction grammars based on artificial data (Doumen et al. 2023), almost always augmented with a pre-defined lexicon (Steels 2004; van Trijp 2008, 2016; Garcia Casademont & Steels 2015, 2016; Garcia Casademont 2018), pre-categorised lexical items (Dominey 2005a, 2005b, 2006; Dominey & Boucher 2005; Beuls et al. 2010; Gerasymova & Spranger 2010, 2012) or a pre-defined set of categories and combinatory rules (Kwiatkowski et al. 2010). Other models in this class learn from natural data augmented with pre-categorised lexical items (Alishahi & Stevenson 2008; Chang 2008) or a predefined set of categories and combinatory rules (Kwiatkowski et al. 2011). A final class of models

learns from natural data enhanced with morpho-syntactic and semantic annotation layers (Dunn 2017, 2018, 2019, 2022, 2023; Dunn & Tayyar Madabushi 2021; Martí et al. 2021). However, these models do not yield grammars that support language comprehension and production.

The third dimension concerns the degree to which the models adhere to the basic tenets of *construction grammar*. Fully constructionist approaches should support bi-directional language processing. They should capture all linguistic knowledge in the form of form-meaning pairings, allow for meaningful schemata above the word level, allow for combining information from different levels of linguistic analysis, and support incremental, individual and adaptive learning. On the most constructionist side of this spectrum, we find approaches that explicitly make use of computational construction grammar implementations to represent and process linguistic structures. Models adopting Fluid Construction Grammar (Steels 2004; van Trijp 2008, 2016; Beuls et al. 2010; Gerasymova & Spranger 2010, 2012; Spranger & Steels 2015; Spranger 2015, 2017; Garcia Casademont & Steels 2015, 2016; Garcia Casademont 2018; Nevens et al. 2022; Doumen et al. 2023) faithfully adhere to the basic tenets of construction grammar. Models adopting  $\mu$ -DOP (Beekhuizen & Bod 2014; Beekhuizen et al. 2014; Beekhuizen 2015) capture form-meaning pairings beyond the level of individual words and support language comprehension and production, but are inherently word order-based. Alishahi & Stevenson (2008), Gaspers et al. (2011, 2016), and Gaspers & Cimiano (2012, 2014) introduce computational operationalisations of construction grammar that are specific for the purposes of the models they describe. In essence, these operationalisations adhere to the basic principles of construction grammar, but they have only been operationalised for basic word order-based argument structure patterns. Models adopting Embodied Construction Grammar (Chang 2008) are constructionist in nature but only support language processing in the comprehension direction. Finally, in the case of models based on Dynamic Construction Grammar (Dominey 2005a, 2005b, 2006; Dominey & Boucher 2005), it is difficult to tell to what extent they capture form-meaning pairings of varying degrees of abstraction, in particular when it comes to modelling phenomena that do not depend on sequential word order. A second class of models makes use of generative grammar formalisms and thereby assumes that the meaning of linguistic expressions mirrors the compositional structure of their syntactic analysis. These models make use of Combinatory Categorical Grammar (Kwiatkowski et al. 2010, 2011, 2012; Artzi & Zettlemoyer 2013; Abend et al. 2017) or probabilistic context-free grammar (Chen & Mooney 2008) as the underlying linguistic framework. Finally, the models by Dunn (2017, 2018, 2019, 2022, 2023), Martí et al. (2021), and Dunn & Tayyar Madabushi (2021) capture partially abstract word

order patterns rather than form-meaning pairings. The slots in these patterns are characterised by either morpho-syntactic or semantic categories.

The final dimension of interest concerns the *operationality* of the resulting grammars. While all models included in this paper are operational in the sense that they have been computationally implemented, not all of them yield grammars that support language comprehension and production. In particular, the models based on ECG (Chang 2008) and CCG (Kwiatkowski et al. 2010, 2011, 2012; Artzi & Zettlemoyer 2013; Abend et al. 2017) only support language comprehension. The models by Dunn (2017, 2018, 2019, 2022, 2023), Martí et al. (2021), and Dunn & Tayyar Madabushi (2021) support neither language comprehension nor production.

A related strand of research that is worth mentioning but falls outside the scope of this paper as defined by the inclusion criteria concerns recent cross-overs between construction grammar and large language models (LLMs). While these approaches share the idea that grammatical patterns can carry non-compositional meaning, they do not (aim to) provide a model for learning construction grammars. As such, Tayyar Madabushi et al. (2020) introduce a methodology for fine-tuning LLMs with the goal of distinguishing between instances of different constructions and Weissweiler et al. (2022, 2023) investigate through probing studies in how far LLMs capture constructional form and meaning. For an overview of work on the intersection of construction grammar and LLMs, we refer the interested reader to Tayyar Madabushi et al. (2025).

From the synthesis presented in this section, we can conclude that almost all challenges involved in computationally learning large-scale, usage-based construction grammars have been addressed in some form in prior research, but that no comprehensive models exist to date. So, what would the ultimate model look like and how can we get there? We hereby put forward a number of milestones that will hopefully serve as a roadmap that can boost and streamline future research efforts.

## Representing meaning

Human languages are constructed by children through interactions with their caregivers. These interactions are meaningful, intentional and situationally grounded. Faithful models of usage-based language acquisition should therefore not rely on direct access to the meaning representations of observed utterances, but implement the process of constructing hypotheses about the intended meaning of the utterances based on the situational context in which they are observed. This entails that the meaning representations need to be composable based on an inventory of pre-linguistic or previously acquired cognitive operations, and



be evaluable with respect to a situational context. This process is referred to as *intention reading* in the psycholinguistic literature (Tomasello 2003). Initial computational operationalisations of this process have been provided by, amongst others, Spranger et al. (2012), Pauw (2013), and Nevens et al. (2022). Semantically annotated corpora are definitely necessary as a scaffold towards the development of large-scale models of construction grammar learning, but cannot be the end point. After all, it is impossible to define, and therefore annotate, meaning independently of contextualised communicative intentions. Ideally, yet probably too ambitious for the short term, the intention reading process should be operationalised using empirically motivated primitive cognitive operations in real-world situations. This will require substantial research efforts into better understanding the building blocks of human cognition, and substantial investments in the development of environments that faithfully simulate the conditions under which human language are acquired, e.g., through the use of virtual reality.

### Representing form

The ultimate model represents utterances in their natural form, i.e., as unsegmented sound waves along with data streams that capture information from other modalities, including eye gaze, facial expression and gesture. This sharply contrasts with how form is handled in current models, where all forms are represented as segmented strings of words, graphemes or phonemes. On the one hand, this entails that the learning of grammars is limited by segmentation choices that have been made upfront and independently of the learning process. On the other hand, these representations leave aside a wealth of potentially meaningful information, for example conveyed through prosody or posture. One important innovation in this direction would be to design algorithms that are able to compare and generalise over unsegmented speech signals on the utterance level. The design of such algorithms could start from prior work on the cross-situational learning of auditory vocabularies from semantically grounded speech data (ten Bosch et al. 2009; Ons et al. 2014; Renkens & Van hamme 2017; Wang & Van hamme 2022) and extend the techniques that are used, so that they can identify patterns above the word level.

### Representing constructions

The ultimate model captures all linguistic knowledge that a language user needs to comprehend and formulate utterances in the form of acquired form-meaning mappings (constructions). This means that no grammar rules, system of categories, or other linguistic structures are predefined. The model can only rely on

general strategies to construct form-meaning mappings, combine them, and generalise over them. The grammar should be able to capture form-meaning mappings of varying degrees of abstraction, so that meaning representations that do not mirror the compositionality of the morpho-syntactic structures that express them can also be modelled. The processing engine that performs construction-based language comprehension and production should be able to combine the information captured in large numbers of constructions, so that morpho-syntactically and/or semantically complex utterances can be handled. Indeed, such utterances typically instantiate a wide variety of constructions. Finally, constructions should be able to incorporate sequential word order patterns, agreement patterns or a combination of both. It is therefore important to use a framework that strongly adheres to these basic principles of construction grammar, such as Fluid Construction Grammar (Steels 2004; van Trijp et al. 2022; Beuls & Van Eecke 2023, 2025).

## Learning constructions

If the input to the language learning process consists of situationally grounded, unsegmented, multi-modal observations of utterances, the first constructions that are learnt can only be holistic pairings between observed utterances and hypotheses about their intended meaning (Tomasello 2003). More general constructions can later be distilled as generalisations over both the form and meaning sides of previously acquired constructions with respect to novel observations. This generalisation process requires access to general syntactico-semantic generalisation algorithms. Initial prototypes of such algorithms have been presented by Van Eecke (2018), Nevens et al. (2022), and Doumen et al. (2023). Apart from improving these algorithms to learn more modular grammars that can, for example, elegantly handle recursive patterns, a crucial target in this direction concerns the design of algorithms that can learn agreement relations on an abstract level, i.e., by expressing congruence without referring to specific categories. Initial efforts in this direction have been presented by Beuls & Höfer (2011), van Trijp & Steels (2012), and Beuls & Steels (2013).

## Language-independent learning

Most prior experiments focus on learning construction grammars based on English data. As a consequence, the resulting models primarily focus on constructions that map between word-order patterns and the meaning they convey. Aspects of meaning expressed through morphology or agreement marking are often not included in the models. While this approach might work for English to a certain

extent, it does not generalise to morphology-rich languages, where much of the function of word order might be taken up by marking strategies. It is important to develop learning mechanisms that are applicable to any language, both from a theoretical and from a practical perspective. Theoretically, it would confirm the constructionist idea that all languages can be modelled through constructions. Practically, it would facilitate the development of construction-based language technology applications for a wide variety of languages.

## Scaling up

A final, more general, criterion concerns the scale of the experiments. As the goal is to learn linguistic capacities from scratch in a human-like manner, it is unavoidable that the language learning process will need to start in very concrete, fully grounded, domain-specific environments. As increasingly more constructions with increasingly higher degrees of abstraction are learnt, the tasks and environments can gradually become more complex, abstract and domain-general. It is important to keep in mind that the ultimate model learns in an incremental fashion, generalises over domains and tasks, and remains forever adaptive to changes in the tasks and environments. While it is natural, and even necessary, to focus on specific parts of the challenge, it is crucial to keep in mind that one day these individual experiments will need to come together.

## 5. Conclusion

The aim of this paper was to provide an overview of prior work concerning computational models of construction grammar learning, to identify gaps in the state of the art and to propose a perspective on the future of the field. We have first described and compared a wide variety of existing models and have then synthesised the state of the art with a special focus on the aim of operationalising usage-based construction grammar on a large scale. Finally, we have formulated a number of milestones that can serve as a roadmap towards the development of scalable, language-independent and adaptive techniques for learning construction grammars in a usage-based fashion.

We have argued that a comprehensive model of construction grammar learning should learn from meaningful, intentional and situationally grounded communicative interactions. As such, meaning representations need to be actively constructed based on the situational context. Form representations should be free from preprocessing artefacts and should not neglect multi-modal information. Bidirectional construction grammars should emerge as a result of applying general

learning strategies and they should cover the full range of linguistic phenomena that occur in the world's languages. We sincerely hope that our synthesis of prior literature and prospective roadmap can help to boost progress in this area of research, streamline efforts undertaken in different research traditions, and bring us closer to language technologies that can learn to use language in a truly natural, human-like manner.

## Funding

The research reported in this paper was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 951846, the Flemish Government under the Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen programme, the Walloon AI Flagship programme ARIAC, and the F.R.S.-FNRS - FWO WEAVE project HERMES I under grant numbers T002724F (F.R.S.-FNRS) and GoAGU24N (FWO). PVE was supported by a postdoctoral fellowship of the Research Foundation Flanders (FWO) with grant No. 75929. JD was funded by the imec.icon project COSMO (HBC.2018.0531) and the imec Smart Education covenant funding Edulab. VJS was funded through a doctoral fellowship awarded by the Research Foundation Flanders (FWO) with grant No. 1108725N.

Open Access publication of this article was funded through a Transformative Agreement with KU Leuven.

## References

-  Abend, O., Kwiatkowski, T., Smith, N.J., Goldwater, S., & Steedman, M. (2017). Bootstrapping language acquisition. *Cognition*, 164, 116–143.
-  Alishahi, A. & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive Science*, 32(5), 789–834.
-  Artzi, Y., & Zettlemoyer, L. (2013). Weakly supervised learning of semantic parsers for mapping instructions to actions. *Transactions of the Association for Computational Linguistics*, 1, 49–62.
- Beekhuizen, B. (2015). *Constructions Emerging* [Doctoral dissertation]. LOT – Netherlands Graduate School of Linguistics.
-  Beekhuizen, B., & Bod, R. (2014). Automating construction work: Data-oriented parsing and constructivist accounts of language acquisition. In R. Boogart, T. Coleman & G. Rutten (Eds.), *Extending the scope of Construction Grammar* (pp. 47–74). Mouton de Gruyter.
-  Beekhuizen, B., Bod, R., Fazly, A., Stevenson, S., & Verhagen, A. (2014). A usage-based model of early grammatical development. In V. Demberg & T. O'Donnell (Eds.), *Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics* (pp. 46–54). Association for Computational Linguistics.
- Bender, E.M. (2008). Grammar engineering for linguistic hypothesis testing. In N. Gaylord, A. Palmer & E. Ponvert (Eds.), *Proceedings of the Texas Linguistics Society X Conference: Computational linguistics for less-studied languages* (pp. 16–36). CSLI.

- Beuls, K., Gerasymova, K., & van Trijp, R. (2010). Situated learning through the use of language games. *Proceedings of the 19th Annual Machine Learning Conference of Belgium and The Netherlands (BeNeLearn)* (pp. 1–6).
- Beuls, K., & Höfer, S. (2011). Simulating the emergence of grammatical agreement in multi-agent language games. In T. Welsh (Ed.), *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* (pp. 61–66). AAAI Press.
-  Beuls, K. & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLOS ONE*, 8(3), e58960.
- Beuls, K. & Van Eecke, P. (2023). Fluid Construction Grammar: State of the art and future outlook. In C. Bonial & H. Tayyar Madabushi (Eds.), *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)* (pp. 41–50). Association for Computational Linguistics.
- Beuls, K. & Van Eecke, P. (2025). Construction grammar and artificial intelligence. In M. Fried & K. Nikiforidou (Eds.), *The Cambridge handbook of Construction Grammar*. Cambridge University Press.
-  Beuls, K., Van Eecke, P., & Cangalovic, V.S. (2021). A computational construction grammar approach to semantic frame extraction. *Linguistics Vanguard*, 7(1), 20180015.
-  Brown, R. (1973). *A first language: The early stages*. Harvard University Press.
-  Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711–733.
- Chang, N. (2008). *Constructing grammar: A computational model of the emergence of early constructions* [Doctoral dissertation]. University of California.
-  Chen, D.L., & Mooney, R.J. (2008). Learning to sportscast: A test of grounded language acquisition. In A. McCallum & S. Roweis (Eds.), *Proceedings of the 25th International Conference on Machine Learning* (pp. 128–135). Association for Computing Machinery.
- Croft, W. (1998). Event structure in argument linking. In M. Butt & W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 21–63). CSLI.
-  Dominey, P.F. (2005a). Emergence of grammatical constructions: Evidence from simulation and grounded agent experiments. *Connection Science*, 17(3–4), 289–306.
-  Dominey, P.F. (2005b). From sensorimotor sequence to grammatical construction: Evidence from simulation and neurophysiology. *Adaptive Behavior*, 13(4), 347–361.
- Dominey, P.F. (2006). From holophrases to abstract grammatical constructions: Insights from simulation studies. In E. Clark & B. Kelly (Eds.), *Constructions in acquisition* (pp. 137–162). CSLI.
-  Dominey, P.F., & Boucher, J.-D. (2005). Learning to talk about events from narrated video in a construction grammar framework. *Artificial Intelligence*, 167(1), 31–61.
-  Doumen, J., Beuls, K., & Van Eecke, P. (2023). Modelling language acquisition through syntactico-semantic pattern finding. In A. Vlachos & I. Augenstein (Eds.), *Findings of the Association for Computational Linguistics: EACL 2023* (pp. 1317–1327). Association for Computational Linguistics.
-  Dunn, J. (2017). Computational learning of construction grammars. *Language and Cognition*, 9(2), 254–292.
- Dunn, J. (2018). Modeling the complexity and descriptive adequacy of construction grammars. *Proceedings of the Society for Computation in Linguistics (SCiL)*, 1, 81–90.

-  Dunn, J. (2019). Frequency vs. association for constraint selection in usage-based construction grammar. In E. Chersoni, N. Hollenstein, C. Jacobs, Y. Oseki, L. Prévot & E. Santus (Eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 117–128). Association for Computational Linguistics.
-  Dunn, J. (2022). Exposure and emergence in usage-based grammar: Computational experiments in 35 languages. *Cognitive Linguistics*, 33(4), 659–699.
- Dunn, J. (2023). Exploring the constructicon: Linguistic analysis of a computational CxG. In C. Bonial & H. Tayyar Madabushi (Eds.), *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)* (pp. 1–11). Association for Computational Linguistics.
-  Dunn, J., & Tayyar Madabushi, H. (2021). Learned construction grammars converge across registers given increased exposure. In A. Bisazza & O. Abend (Eds.), *Proceedings of the 25th Conference on Computational Natural Language Learning* (pp. 268–278). Association for Computational Linguistics.
- EHAI (2023). SemBrowse: Semantics-driven corpus exploration. <https://ehai.ai.vub.ac.be/sembrowse/>
-  Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1), 1–24.
-  Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34(6), 1017–1063.
- Garcia Casademont, E. (2018). *Origins of recursive phrase structure through cultural self-organisation and selection* [Doctoral dissertation]. Universitat Pompeu Fabra.
- Garcia Casademont, E., & Steels, L. (2015). Usage-based grammar learning as insight problem solving. In G. Airenti, B. G. Bara & G. Sandini (Eds.), *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science* (pp. 258–263). CEUR Workshop Proceedings.
-  Garcia Casademont, E., & Steels, L. (2016). Insight grammar learning. *Journal of Cognitive Science*, 17(1), 27–62.
-  Gaspers, J., & Cimiano, P. (2012). A usage-based model for the online induction of constructions from phoneme sequences. *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 1–6.
-  Gaspers, J., & Cimiano, P. (2014). A computational model for the item-based induction of construction networks. *Cognitive Science*, 38(3), 439–88.
-  Gaspers, J., Cimiano, P., Griffiths, S. S., & Wrede, B. (2011). An unsupervised algorithm for the induction of constructions. *2011 IEEE International Conference on Development and Learning (ICDL)*, 1–6.
-  Gaspers, J., Cimiano, P., Rohlfing, K., & Wrede, B. (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.
- Gerasymova, K., & Spranger, M. (2010). Acquisition of grammar in autonomous artificial systems. In H. Coelho, R. Studer & M. Woolridge (Eds.), *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI-2010)* (pp. 923–928). IOS Press.
-  Gerasymova, K., & Spranger, M. (2012). An experiment in temporal language learning. In L. Steels & M. Hild (Eds.), *Language grounding in robots* (pp. 237–254). Springer.

- Goldberg, A.E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219–224.
- Hemphill, C.T., Godfrey, J.J., & Doddington, G.R. (1990). The ATIS spoken language systems pilot corpus. *Speech and Natural Language: Proceedings of a Workshop held at Hidden Valley* (pp. 96–101).
- Johnson, J., Hariharan, B., van der Maaten, L., Fei-Fei, L., Lawrence Zitnick, C., & Girshick, R. (2017). CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1988–1997). IEEE.
- Krenn, B., Sadeghi, S., Neubarth, F., Gross, S., Trapp, M., & Scheutz, M. (2020). Models of cross-situational and crossmodal word learning in task-oriented scenarios. *IEEE Transactions on Cognitive and Developmental Systems*, 12(3), 658–668.
- Kwiatkowski, T., Goldwater, S., Zettlemoyer, L., & Steedman, M. (2012). A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their meanings. In W. Daelemans (Ed.), *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 234–244). Association for Computational Linguistics.
- Kwiatkowski, T., Zettlemoyer, L., Goldwater, S., & Steedman, M. (2010). Inducing probabilistic CCG grammars from logical form with higher-order unification. In H. Li & L. Màrquez (Eds.), *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 1223–1233). Association for Computational Linguistics.
- Kwiatkowski, T., Zettlemoyer, L., Goldwater, S., & Steedman, M. (2011). Lexical generalization in CCG grammar induction for semantic parsing. In R. Barzilay & M. Johnson (Eds.), *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 1512–1523). Association for Computational Linguistics.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk (3rd edition)*. Lawrence Erlbaum Associates.
- Martí, M.A., Taulé, M., Kovatchev, V., & Salamó, M. (2021). DISCOVer: DIStributIonal approach based on syntactic dependencies for discovering CONstructions. *Corpus Linguistics and Linguistic Theory*, 17(2), 491–523.
- Müller, S. (2015). The coregram project: Theoretical linguistics, theory development, and verification. *Journal of Language Modelling*, 3(1), 21–86.
- Nevens, J., Doumen, J., Van Eecke, P., & Beuls, K. (2022). Language acquisition through intention reading and pattern finding. In N. Calzolari & C.-R. Huang (Eds.), *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 15–25). International Committee on Computational Linguistics.
- Ons, B., Gemmeke, J.F., & Van hamme, H. (2014). Fast vocabulary acquisition in an NMF-based self-learning vocal user interface. *Computer Speech & Language*, 28(4), 997–1017.
- Pauw, S. (2013). *Size matters: Grounding quantifiers in spatial perception* [Doctoral dissertation]. University of Amsterdam.
- Renkens, V., & Van hamme, H. (2017). Automatic relevance determination for nonnegative dictionary learning in the gamma-poisson model. *Signal Processing*, 132, 121–133.
- Spranger, M. (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)* (pp. 196–201). IEEE.

- Spranger, M. (2017). Usage-based grounded construction learning: A computational model. In *The 2017 AAAI Spring Symposium Series* [Technical report] (pp. 245–250). AAAI Press.
-  Spranger, M., Pauw, S., Loetzsch, M., & Steels, L. (2012). Open-ended procedural semantics. In L. Steels, & M. Hild (Eds.), *Language grounding in robots* (pp. 153–172). Springer.
- Spranger, M., & Steels, L. (2015). Co-acquisition of syntax and semantics: an investigation in spatial language. In Q. Yang, & M. Wooldridge (Eds.), *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence* (pp. 1909–1915). AAAI Press.
-  Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103(1–2), 133–156.
-  Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent Systems*, 16, 16–22.
-  Steels, L. (2004). Constructivist development of grounded construction grammar. *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)* (pp. 9–16).
-  Steels, L., & De Beule, J. (2006). Unify and merge in Fluid Construction Grammar. In P. Vogt, Y. Sugita, E. Tuci & C. L. Nehaniv (Eds), *Symbol grounding and beyond, International Workshop on Emergence and Evolution of Linguistic Communication (EELC 2006)* (pp. 197–223). Springer.
-  Tayyar Madabushi, H., Romain, L., Divjak, D., & Milin, P. (2020). CxGBERT: BERT meets Construction Grammar. In D. Scott, N. Bel & C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 4020–4032). International Committee on Computational Linguistics.
- Tayyar Madabushi, H., Romain, L., Milin, P., and Divjak, D. (2025). Construction Grammar and language models. In M. Fried, & K. Nikiforidou (Eds.), *The Cambridge handbook of Construction Grammar*. Cambridge University Press.
-  ten Bosch, L., Boves, L., Van hamme, H., & Moore, R. K. (2009). A computational model of language acquisition: The emergence of words. *Fundamenta Informaticae*, 90(3), 229–249.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Van Eecke, P. (2018). *Generalisation and specialisation operators for computational construction grammar and their application in evolutionary linguistics research* [Doctoral dissertation]. Vrije Universiteit Brussel, VUB Press.
- van Trijp, R. (2008). *Analogy and multi-level selection in the formation of a Case Grammar. A case study in Fluid Construction Grammar* [Doctoral dissertation]. University of Antwerp.
-  van Trijp, R. (2016). *The evolution of case grammar*. Language Science Press.
-  van Trijp, R., Beuls, K., & Van Eecke, P. (2022). The FCG Editor: An innovative environment for engineering computational construction grammars. *PLOS ONE*, 17(6), e0269708.
-  van Trijp, R., & Steels, L. (2012). Multilevel alignment maintains language systematicity. *Advances in Complex Systems*, 15(3–4), 1250039.
-  Verheyen, L., Botoko Ekila, J., Nevens, J., Van Eecke, P., & Beuls, K. (2023). Neuro-symbolic procedural semantics for reasoning-intensive visual dialogue tasks. In K. Gal, A. Nowé, G. J. Nalepa, R. Fairstein, & R. Rădulescu (Eds.), *Proceedings of the 26th European Conference on Artificial Intelligence (ECAI 2023)* (pp. 2419–2426). IOS Press.
-  Wang, P., & Van hamme, H. (2022). Bottleneck low-rank transformers for low-resource spoken language understanding. *Interspeech 2022*, 1248–1252.



- Weissweiler, L., He, T., Otani, N., R. Mortensen, D., Levin, L., & Schütze, H. (2023). Construction grammar provides unique insight into neural language models. In C. Bonial, & H. Tayyar Madabushi (Eds.), *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)* (pp. 85–95). Association for Computational Linguistics.
-  Weissweiler, L., Hofmann, V., Köksal, A., & Schütze, H. (2022). The better your syntax, the better your semantics? Probing pretrained language models for the English comparative correlative. In Y. Goldberg, Z. Kozareva & Y. Zhang (Eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 10859–10882). Association for Computational Linguistics.
-  Willaert, T., Van Eecke, P., Beuls, K., & Steels, L. (2020). Building social media observatories for monitoring online opinion dynamics. *Social Media + Society*, 6(2).
- Zelle, J. M., & Mooney, R. J. (1996). Learning to parse database queries using inductive logic programming. *Proceedings of the Thirteenth National Conference on Artificial Intelligence – Volume 2* (pp. 1050–1055).

## Address for correspondence

Paul Van Eecke  
 Vrije Universiteit Brussel (DINF-ARTI)  
 Pleinlaan 2  
 B-1050 Brussel  
 Belgium  
 paul@ai.vub.ac.be  
 <https://orcid.org/0000-0001-9153-9092>

## Co-author information

Jonas Doumen  
 Department of Linguistics, Faculty of Arts,  
 KU Leuven, Leuven, Belgium  
 Itec, Imec research group at KU Leuven,  
 Kortrijk, Belgium  
[jonas.doumen@kuleuven.be](mailto:jonas.doumen@kuleuven.be)

Veronica Juliana Schmalz  
 Department of Linguistics, Faculty of Arts,  
 KU Leuven, Leuven, Belgium  
 Itec, Imec research group at KU Leuven,  
 Kortrijk, Belgium  
[veronicajuliana.schmalz@kuleuven.be](mailto:veronicajuliana.schmalz@kuleuven.be)

Katrien Beuls  
 Faculté d'informatique  
 Université de Namur  
[katrien.beuls@unamur.be](mailto:katrien.beuls@unamur.be)

## Publication history

Date received: 2 June 2023

Date accepted: 20 February 2024

Published online: 16 December 2024