Constructions All the Way Up: From Sensory Experiences to Construction Grammars

Jérôme Botoko Ekila*
Artificial Intelligence Laboratory
Vrije Universiteit Brussel, Belgium
jerome@ai.vub.ac.be

Lara Verheyen*
Artificial Intelligence Laboratory
Vrije Universiteit Brussel, Belgium
lara.verheyen@ai.vub.ac.be

Katrien Beuls†
Faculté d'informatique
Université de Namur, Belgium
katrien.beuls@unamur.be

Paul Van Eecke†
Artificial Intelligence Laboratory
Vrije Universiteit Brussel, Belgium
paul@ai.vub.ac.be

Abstract

Constructionist approaches to language posit that all linguistic knowledge is captured in constructions. These constructions pair form and meaning at varying levels of abstraction, ranging from purely substantive to fully abstract and are all acquired through situated communicative interactions. In this paper we provide computational support for these foundational principles. We present a model that enables an artificial learner agent to acquire a construction grammar directly from its sensory experience. The grammar is built from the ground up, i.e. without a given lexicon, predefined categories or ontology and covers a range of constructions, spanning from purely substantive to partially schematic. Our approach integrates two previously separate but related experiments, allowing the learner to incrementally build a linguistic inventory that solves a question-answering task in a synthetic environment. These findings demonstrate that linguistic knowledge at different levels can be mechanistically acquired from experience.

1 Introduction

According to constructionist approaches to language (Fillmore, 1988; Goldberg, 1995; Croft, 2001; Goldberg, 2003) all linguistic knowledge is captured in constructions, pairing form and meaning. Within this framework, constructions vary in their level of abstraction, ranging from purely substantive to fully abstract, all shaped by usage. As Goldberg (2003, p. 223) famously put it: *"it's constructions all the way down"*.

Constructions are not abstract templates shared uniformly between members of a linguistic community, rather each one is grounded in an individual's embodied experience and interaction with the world (Lakoff, 1987; Langacker, 1987; Bybee, 2010; Tomasello, 2003; Diessel, 2017). For instance, a construction mapping the form *"dog"* to its underlying DOG concept is shaped by an individual's encounters with dogs, including what they have seen, learned or heard about them. Beyond the perceptual level, language users also acquire constructions that coordinate more abstract cognitive processes (Goldberg, 1995). Consider the sentence *"The dog chases the cat."* in which the transitive construction organises the relation between a CHASING event and its participants. This abstract relation is learned through repeated encounters of linguistic utterances and observations in the world. Whether the meaning of a construction is a concept grounded in direct sensory experience or an abstract schema, all are pairings of form and meaning and arise from situated interactions (Beuls and Van Eecke, 2025). This linguistic knowledge is built up through cognitive mechanisms that reconstruct the intended meaning of an interlocutor and find patterns over form-meaning mappings (Tomasello, 2003; Dąbrowska and Lieven, 2005; Behrens, 2009; Lieven, 2014).

A computational approach to modelling language acquisition involves language games, in which embodied agents acquire constructions through repeated situated communicative interactions (Steels, 1995, 1999). These simulations offer a mechanistic model of language acquisition, and have been used to study the emergence of linguistic structure at multiple levels, from basic grounded

*Joint first authors.
†Joint last authors.

Proceedings of the Second International Workshop on Construction Grammars and NLP, pages 84–95
September 24, 2025,

lexicons (Steels, 1995; Kaplan et al., 1998; Loetzsch, 2015; Nevens et al., 2020; Botoko Ekila et al., 2024) to early forms of syntax (De Beule and Bergen, 2006; Van Eecke, 2018) and more complex grammatical systems (van Trijp, 2008; Beuls and Höfer, 2011; Spranger and Steels, 2015; Steels and Garcia Casademont, 2015; Nevens et al., 2022; Doumen et al., 2024). However, a key challenge remains unsolved: no existing computational model has yet demonstrated the emergence of a construction grammar that is both directly learned from sensory experience and capable of capturing a range of constructions, spanning from fully substantive constructions to more abstract constructions, without a given lexicon, ontology or predefined categories.

In this paper, we present a model that enables a learner agent to acquire a construction grammar from the ground up through situated communicative interactions with a tutor agent. Using a curriculum learning approach, where training progresses from simpler to more complex interactions, the learner develops a grammar that spans from perceptually grounded lexical constructions to partially schematic constructions. We validate our approach experimentally in a synthetic continuous environment in which a learner develops a grammar to interpret and answer questions. We thereby demonstrate that, with the help of a tutor, a computational construction grammar including more abstract constructions can be acquired directly from sensory experience, supporting the hypothesis that it is also, indeed, *constructions all the way up*.

2 Background

The model we present is embedded within the framework of language games (Steels, 1995, 1999), which is used to simulate how agents can establish linguistic conventions through repeated situated communicative interactions. In this paper, we build on language acquisition experiments that each focus on different levels of abstraction: (i) acquiring perceptually grounded lexical constructions that link sensory experiences to linguistic forms (Nevens et al., 2020; Botoko Ekila et al., 2024) and (ii) acquiring grammatical constructions that capture structural patterns in language use (Nevens et al., 2022; Doumen et al., 2024). Although these four experiments focus on acquiring constructions at varying levels of abstraction, they rely on the same shared principle: agents acquire form-meaning mappings through situated communicative interactions (Beuls and Van Eecke, 2024). The next sections summarise the core mechanisms behind each experiment, which forms the basis of our integrated approach.

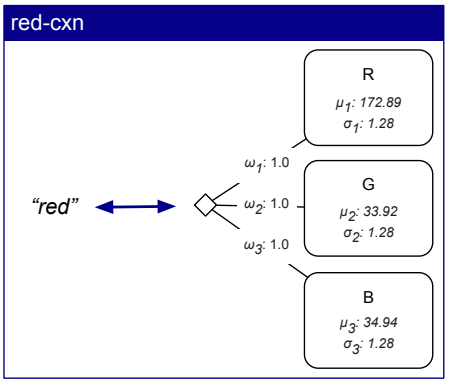

Figure 1: Example of a grounded lexical construction learned by the learner agent for the word *"red"*, which has specialised towards the three colour feature dimensions (RGB). Only dimensions with weights greater than 0.0 are shown.

2.1 Acquiring grounded lexical constructions

The experiments of Nevens et al. (2020) and Botoko Ekila et al. (2024) are concerned with acquiring form-meaning pairings that link sensory experiences to linguistic forms. In this process, a learner agent acquires a set of constructions that capture perceptual concepts such as RED or LARGE by interacting with a tutor agent. Importantly, the learner starts without any prior linguistic knowledge.

In the experiments, both agents are situated in a shared environment with different objects and engage in a series of referential games, each corresponding to a single interaction. In each interaction, the tutor (i) selects a target object from the scene and (ii) produces a single-word utterance that refers to a property of the selected object that distinguishes it from the other objects. The learner observes the scene through its own sensors, which capture raw perceptual features (e.g. RGB for colour or the number of pixels an object occupies in the image for size). The goal of the learner is to infer which object the tutor is referring to, based on the utterance, the perceptual input, and any linguistic knowledge acquired in previous interactions. After each interaction, the tutor reveals the correct referent (i.e. the target object), providing explicit feedback. At no point are the tutor's and learner's internal representations shared between agents. The learner must refine its own internal

representations through these interactions with the tutor. They store the observed word forms (e.g. *"red"*) and associated internal concept representations as form-meaning mappings in its inventory.

Concepts are modelled as weighted Gaussian distributions over sensory features. Each distribution captures the prototypical range of values associated with that feature, while the associated weight captures the feature's relevance to the concept. For example, as seen in Figure 1, the concept linked to the word *"red"* assigns high weights to RGB features and low weights to other features. These distributions and weights are updated incrementally through repeated interactions with the tutor.

Early on, the learner's answers are mostly incorrect, but as they interact more, the learner refines its concept representations based on the feedback of the tutor. Over time, the learner builds a conceptual system grounded in its own sensory experience of the world.

2.2 Acquiring grammatical constructions

In the experiments of Nevens et al. (2022) and Doumen et al. (2024), a learner acquires lexical and grammatical constructions by playing a question-answering game. The game operates in a symbolic representation of the environment of the experiments discussed in Section 2.1. In this symbolic version of the setting, objects are described using structured attribute-value pairs (e.g. OBJECT-1: {COLOUR: RED, SHAPE: CUBE}). This setup abstracts away from raw sensory inputs and perceptual processing, allowing the learner to work directly with high-level representations of objects. Thus, as seen in Figure 2, the meaning of the CUBES-CXN is represented by the symbol CUBE.

Within this symbolic setting, the tutor poses questions about a scene such as *"How many red cubes are there?"* or *"What shape does the blue object have?"*. The learner's task is to interpret the question and produce a correct answer. To achieve this, the learner builds a construction grammar that maps linguistic utterances to meaning representations that can be executed to retrieve the answer. To acquire these constructions, the learner is equipped with two core learning mechanisms: *intention reading* and *pattern finding* (Tomasello, 2003). Intention reading refers to a language user's ability to reconstruct the intended meaning of an utterance, enabling the learner to hypothesise about the speaker's intended meaning. Pattern finding

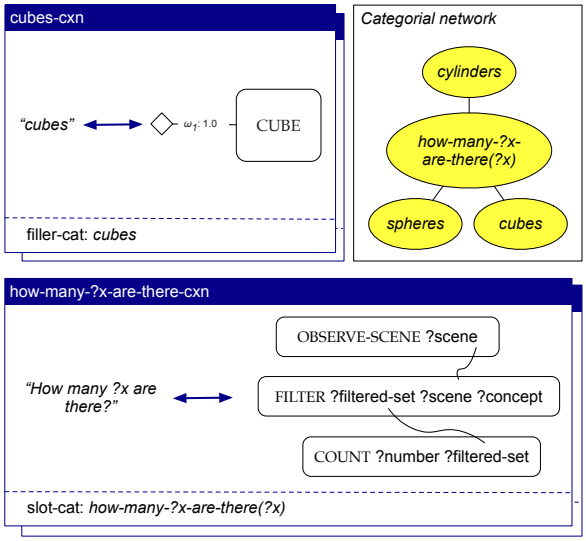

Figure 2: Example of two constructions acquired by the learner agent during the question-answering game that takes place in a symbolic environment. A lexical CUBES-CXN with a symbolic concept representation, an item-based HOW-MANY-?X-ARE-THERE-CXN and part of the categorial network, capturing the slot-filler relation through the categories in the constructions, are shown. Figure based on Nevens et al. (2022) and Doumen et al. (2024).

refers to the ability to generalise across different communicative interactions. We briefly summarise how these processes are operationalised, but for a more comprehensive explanation we refer the reader to Nevens et al. (2022) and Doumen et al. (2024).

The learner starts the game with an empty linguistic inventory but is endowed with a set of atomic cognitive operations (so-called *primitive operations*). The meaning of questions is represented as sequences of these operations, each of which are needed to find the correct answer, i.e. a form of procedural semantics (Winograd, 1972; Woods, 1968). Formally, each question is encoded as a set of predicates. Each predicate corresponds to a *primitive operation* that the learner can perform, such as filtering objects by their properties or counting elements in a set. For example, the question *"How many cubes are there?"* can be represented as a sequence of three primitive operations: (i) observing the current scene with OBSERVE-SCENE, (ii) filtering for objects of type cube with FILTER, and (iii) counting the resulting set with COUNT.

At the start of each interaction, both agents are situated in the same scene. The tutor then poses a question to the learner about the scene. The learner

attempts to interpret and answer the question using its current linguistic inventory. If the learner fails to interpret the question or the answer is incorrect, the tutor provides feedback in the form of the correct answer. The learner then attempts to recover the intended meaning by abductively reasoning about the tutor's communicative goal (i.e. *intention reading*). In doing so, it searches for a program (a sequence of primitive operations) that would lead to the tutor's answer. Once a plausible program is found, the learner can store this new utterance-program pairing as a candidate construction.

Over time, through an inductive process, the learner generalises across observed utterances and reconstructed meanings to build more abstract schemata (i.e. *pattern finding*). For example, if the learner has previously encountered and understood the question *"How many spheres are there?"* and then observes *"How many cubes are there?"*, it can induce a pattern. As shown in Figure 2, one possible generalisation could yield a construction that includes a slot, e.g. HOW-MANY-?X-ARE-THERE?-CXN, and another that can fill that slot, e.g. CUBES-CXN. A construction can thus be partially schematic: containing both fixed elements and variable slots. Slots are the parts that remain open and available to be filled by other constructions. Constructions may contain more than one slot, and slots can also occur adjacently. In the remainder of this text, we refer to partially schematic constructions with one or more slots as *item-based* constructions, while fully substantive constructions are referred to as *lexical* constructions.

As the construction inventory grows, the learner becomes able to interpret parts of novel utterances. The learner can then use this partial analysis as a starting point to more efficiently search for the remaining operations needed to construct a full program that leads to the answer. In total, seven generalisation operators are introduced by Nevens et al. (2022) and Doumen et al. (2024).

A critical component of the approach is the *categorial network* which organises the learner's acquired knowledge of which constructions can fill in slots of other constructions (Van Eecke, 2018). As seen in Figure 2, the *how-many-?x-are-there(?x)* category is linked to three filler categories (*spheres*, *cylinders*, *cubes*) that can fill the *?x* slot. The categorial network thus stores slot-filler relations observed during interactions and dynamically expands as new combinations are encountered. This mechanism supports an important generalisation: even when the learner has never seen a particular combination of constructions, it can still interpret the utterance if the individual components are known. For example, the learner might already know a construction WHAT-IS-THE-?X-MADE-OF?-CXN and another SPHERE-CXN, but never observed the specific combination *"What is the sphere made of?"*. In such cases, the categorial network allows the learner to combine known constructions by creating a new link between these categories, without needing to create a new construction.

Together, intention reading, pattern finding and the categorial network form the core mechanisms through which the learner agent acquires a flexible and compositional grammar. Through this grammar, the agent can solve the task of interpreting and answering the questions.

3 Acquiring a Construction Grammar from Sensory Experience

To demonstrate how a computational construction grammar spanning multiple levels of abstraction can be acquired directly from sensory experience, we integrate the experiments discussed in Sections 2.1 and 2.2. In our integrated methodology, a learner agent first acquires grounded lexical constructions through a reference-based game, before using these constructions as building blocks in a question-answering game.

To achieve this integration, the symbolic scene representations used in the question-answering game must be replaced by a continuous environment. As discussed in Section 2.2, the original experiment assumes symbolic input in the form of structured representations. This allows the learner to reason directly over discrete, high-level structures using its primitive operators, bypassing the challenge of perceptual grounding. In the continuous setting, the primitive operators must be adapted to reason over low-level structures which is not straightforward. We highlight three key changes.

Similarity between concepts and objects A crucial cognitive operation in the experiment is the FILTER primitive, which takes a set of objects as input and returns a filtered set containing only those objects for which a given concept applies. In the original symbolic setting, filtering objects by a concept relied on symbolic matching. In our continuous setup, we adapt the FILTER primitive to work

with raw perceptual features. Rather than checking whether an object has a given symbolic feature, the learner now computes a similarity score between the grounded concept and each object in the input set. This similarity is calculated using the algorithm introduced in Nevens et al. (2020). It estimates the likelihood that an object's sensory features were generated by the concept's distribution. Any object whose similarity exceeds a threshold γ is included in the filtered set.

Deriving category hierarchies The original question-answering experiments operate under a critical assumption: agents must also already have a prespecified *category hierarchy* (Rosch et al., 1976) to perform certain basic cognitive operations, such as querying on a category (e.g. COLOUR). The learner is assumed to already understand, for example, that SIZE constitutes a superordinate category with mutually exclusive values like SMALL and LARGE. This assumption provides a scaffold that simplifies the problem, but it does raise questions about how such hierarchies can be acquired.

It has been hypothesised that a categorial network, capturing slot-filler relationships, contains the information needed to derive these category hierarchies (Van Eecke, 2018; Steels et al., 2022; Nevens et al., 2022; Doumen et al., 2024). Simply put, categories that behave similarly across constructions may belong to the same domain. To operationalise this hypothesis, we identify potential semantic fields: groups of categories that likely belong to the same domain. This is achieved by clustering categories based on their *constructional behaviour* captured by the categorial network. Concretely, we compute the vertex cosine similarity (Salton and McGill, 1983) between all pairs of categories (i.e. fillers) in the categorial network. This yields a fully connected graph where each node corresponds to a category and each edge is weighted by the similarity score. Categories that frequently fill the same slots will have many shared connections, and thus a higher vertex cosine similarity. To identify meaningful clusters, we apply a threshold τ and retain only edges with high similarity scores. This pruning step breaks the network into connected components that represent potential semantic fields, such as size, colour or shape.

Generalisation operators As discussed in Section 2.2, the original question-answering experiment uses seven generalisation operators. These

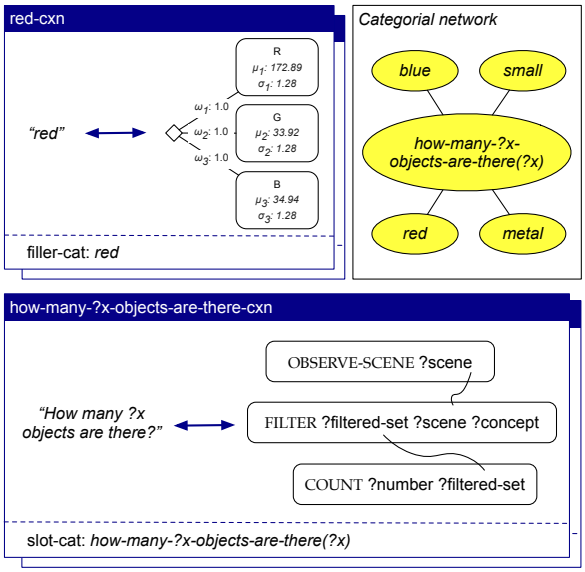

Figure 3: Example of two constructions acquired by the learner during the question-answering game in the continuous setting. A lexical RED-CXN, an item-based HOW-MANY-?X-OBJECTS-ARE-THERE-CXN and part of the categorial network, capturing the slot-filler relation through the categories in the constructions, are shown.

included generalisations over holistic mappings between linguistic forms and reconstructed meanings. These types of generalisations could lead to item-based and lexical constructions. In our experiment, all required lexical constructions are acquired before the question-answering game begins. As a result, operators that generalise over holistic mappings that yield lexical constructions are no longer needed. Due to our two-phased approach, only three generalisation operators are used (i.e. *add-categorial-link*, *lexical→item-based* and *nothing→holophrase*).

4 Experimental Validation

We validate our methodology experimentally. The experiment is structured in two phases. Initially, the learner acquires concepts through a reference-based game using the methodology discussed in Section 2.1. After this phase, the learner has acquired a set of grounded lexical constructions that are mappings between linguistic forms and perceptually grounded concept representations. In the second phase, the learner participates in a question-answering game using the adaptations discussed in Section 3. Concretely, the learner agent further expands its construction inventory with item-based constructions, in which the previously acquired

grounded lexical constructions serve as fillers. Figure 3 captures this idea: the construction inventory of the agent consists of both grounded lexical constructions as well as item-based constructions linked through the categorial network. In contrast to Figure 2, the meaning of the lexical construction is now represented by a grounded concept. Importantly, although the two phases are structured sequentially, learning is not confined to each phase: during the second phase, the concept representations in the grounded lexical constructions continue to be refined through new observations.

Data The experiment uses the CLEVR dataset (Johnson et al., 2017). This dataset contains questions about images containing three to ten geometric objects. Each object is described by a combination of attributes: one of three shapes (SPHERE, CUBE or CYLINDER), one of eight colours (GREY, BLUE, BROWN, YELLOW, RED, GREEN, PURPLE or CYAN), one of two material types (METAL or RUBBER) and one of two sizes (SMALL or LARGE).

Following Nevens (2022) and Doumen et al. (2024), we use a subset of the CLEVR scenes and questions. To create the continuous environment, we extract features for each object in the scenes from the dataset following the data processing steps discussed in Nevens et al. (2020, p. 7). We use 14,000 of the 15,000 scenes across both *Phase 1* and *2* and hold out the remaining 1,000 scenes for evaluation. This allows us to assess how well the methodology generalises to previously unseen scenes after *Phase 2*. Only questions involving (i) counting, (ii) checking for existence and (iii) querying for a certain attribute are retained. To obtain this subset, we removed questions related to comparison, spatial relations and logical operations. As explained in Doumen et al. (2024), this choice is motivated by the complexity of these operations which is far removed from the complexity of the questions that children encounter in the beginning of the language acquisition process. Lastly, in the CLEVR dataset, synonyms are used to describe the exact same concepts (e.g. sphere and ball). We remove these questions, following the principle of no synonymy (Goldberg, 1995, p. 67). Thus, in *Phase 2* of the experiment 1,935 unique questions can be posed about 14,000 different scenes.

Experimental setup The experimental setup of *Phase 1* follows Nevens et al. (2020). *Phase 2*, due to its increased complexity, is further broken down into three successive steps to facilitate learning. First, the tutor poses counting and existence check questions (respectively named *Phase 2A* and *Phase 2B*). This allows the learner to observe many slot-filler relationships and gradually build up its categorial network. After this, the tutor moves onto questions related to querying attributes of objects (*Phase 2C*), which requires reasoning over category hierarchies. These hierarchies are created based on the categorial network that was built up during the previous phases using the category clustering method described in Section 3. The thresholds γ and τ are hyperparameters and are set empirically to respectively $\gamma = 0.8$ and $\tau = 0.7$. In total the experiment consists of 20,000 interactions. *Phase 1* consists of 5,000 interactions, while *Phase 2* consists of 5,000 interactions for each of the three parts: *Phase 2A*, *2B* and *2C*. All reported results are averaged over 10 independent runs. All runs were conducted on a 12-core CPU paired with 16GB of RAM, with each run completed in ± 0.5 hours.

Learning dynamics The learning dynamics of the experiment are shown in Figure 4. For each phase, we keep track of the average communicative success over time. For the reference-based game, an interaction is successful if the learner points to the tutor's intended referent. For the question-answering game, there is success when the learner utters the expected answer. In both cases, a success of 100% means that learner understands the tutor perfectly.

As seen in Figure 4, at the end of *Phase 1*, an inventory of 15 grounded lexical constructions is acquired.[1] In the beginning of *Phase 2A*, when the learner encounters questions related to counting, success drops down, but quickly rises again when the learner successfully acquires item-based constructions that are needed to answer the questions. By the end of this phase, on average, 20 item-based constructions and 4 holophrase constructions are acquired and the necessary links between the slots of the item-based constructions and the slot-filler relations are learned and added to the categorial network. Similar dynamics are observed when the existence and query questions are introduced (*Phases 2B and 2C*). First, the success drops down, but the

[1] Note that the number of lexical constructions jumps from 15 to 18 between *Phase 1* and *Phase 2*. This increase is due to the creation of three plural equivalents for the singular 'shape'-constructions. In *Phase 1*, the tutor only refers to singular concepts, but later in the experiment, the plural versions are required.

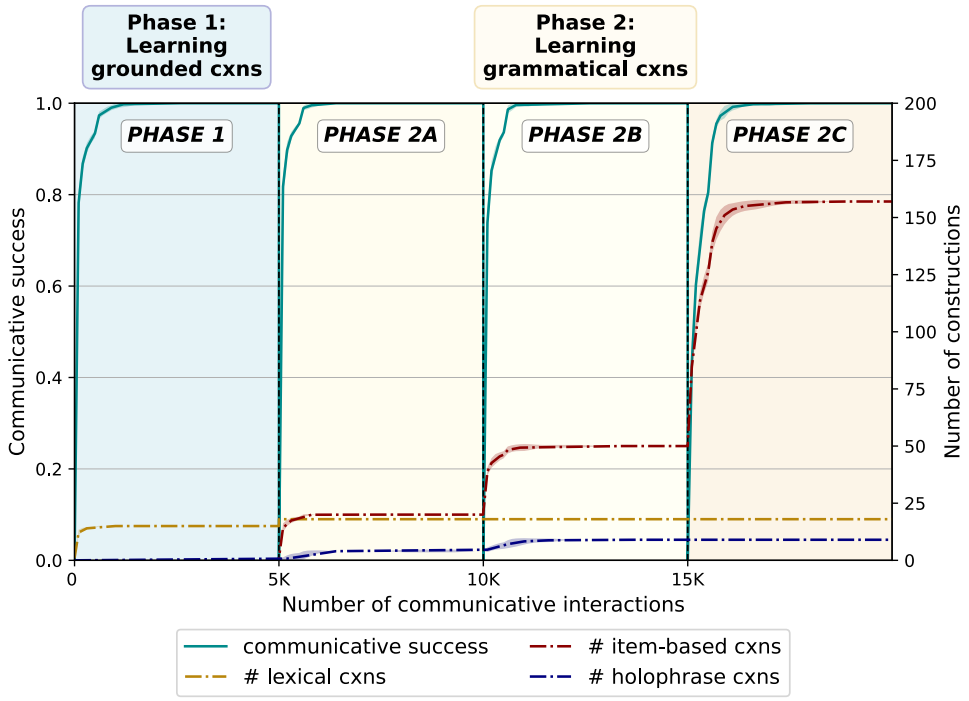

Figure 4: Learning dynamics of the experiment. The blue line denotes the degree of average communicative success over the past 500 interactions. At the start of each phase the average window is reset. The other dashed lines (yellow, red and blue) respectively denote the number of lexical, item-based and holophrase constructions that were acquired over time. At the beginning of each phase, communicative success drops down, but quickly recovers as constructions are acquired to resolve communicative impasses. Results are averaged over 10 independent runs.

agent quickly acquires the necessary constructions and communicative success is reached again after a couple of hundred interactions. The linguistic inventory size expands to $\pm$ 50 item-based and 9 holophrase constructions at the end of *Phase 2B* and reaches a number of 157 item-based constructions at the end of the experiment, leading to a total of $\pm$ 184 constructions.[2] Finally, we evaluate the acquired construction grammar on a held-out set of 1,000 unseen scenes in ten independent runs. During this phase, the learner's linguistic system is frozen and cannot be updated. We perform 5,000 additional interactions on this evaluation set. The learner correctly interprets and answers the tutor's question posed in 99.65% of interactions, averaged over 10 independent runs. Analysis of the rare failure cases reveals that errors are primarily due to grounding issues, where slight out-of-distribution observations relative to the learned concept representations lead down the line to incorrect answers.

Formation of a category hierarchy The categorial network captures the slot-filler relations of the constructions. These relations are built up during the experiment and form the basis for the formation of category hierarchies, which are used in the last phase of the experiment.

Figure 5 shows the expansion of the learner's categorial network during the different phases of the experiment. For visual purposes, we zoom in on categories related to nine grounded lexical constructions and three item-based constructions. After *Phase 1*, the network consists only of disconnected categories for grounded lexical constructions. During *Phase 2A* categories start to cluster. We observe that categories that relate to the shape of objects act as fillers in similar slots (e.g. they fill the *?y* slot in the HOW-MANY-?X-?Y-ARE-THERE-CXN) and are thus possibly more related to each other than, for example, the 'material', 'colour' and 'size' categories which fill the *?x* slot in the same construction (see Figure 5). During *Phase 2B*, the categorial network expands. Now, we clearly ob-

[2]Note that there is no typical 'overshoot' pattern for the number of constructions. In the reference-based game, this is due to the lack of ambiguity regarding form about which forms map to which meaning. The learner directly acquires a construction with an initial concept representation that is gradually refined. In the question-answering game, we observe that the agent likewise acquires the optimal meaning representation from the start. This contrasts with the original experiment, where many suboptimal lexical mappings were first acquired. Our two-phased approach prevents lexical suboptimal hypotheses.

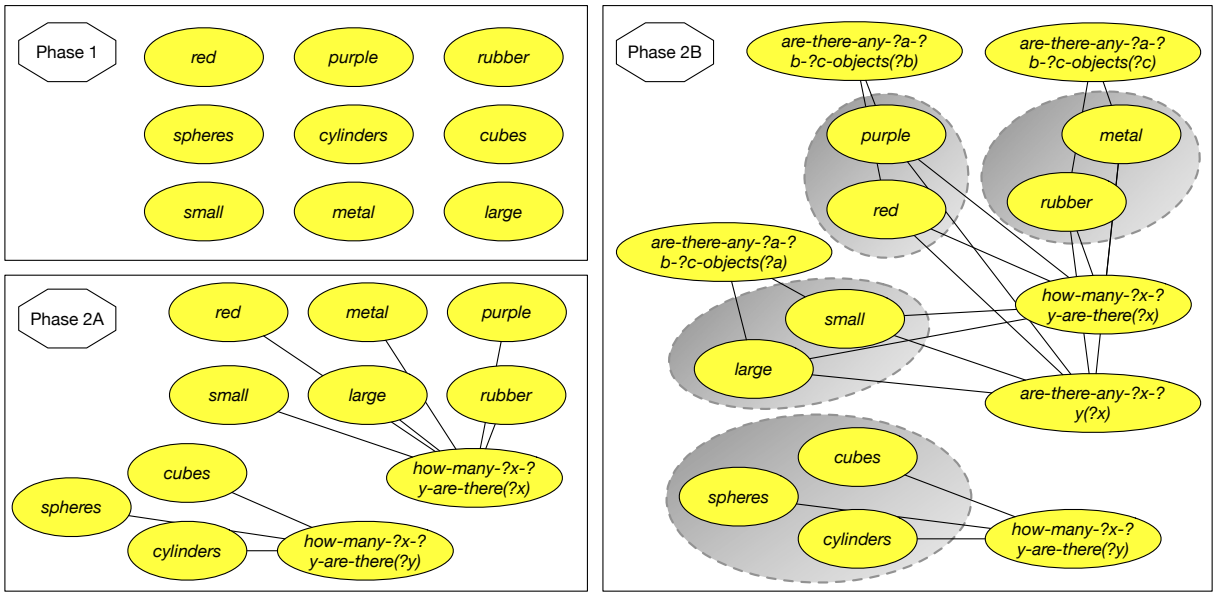

Figure 5: Expansion of the learner's categorial network over the course of the experiment. *Phase 1* shows the network after the grounded lexical construction learning phase, with no links between categories yet. In *Phase 2A* initial clusters of categories begin to form. In *Phase 2B*, shaded in grey, four semantically relevant clusters emerge (size, colour, material, shape). Only a subset of the categorial network is shown for illustrative purposes.

serve that categories cluster together into meaningful hierarchies over concepts, indicated by the grey shaded regions in Figure 5. This cluster formation is used to build the hierarchy needed in *Phase 2C* of the experiment in which the agent needs to query attributes of certain objects. Applying the methodology described in Section 3 results in 5 clusters: the shapes (singular and plural), the colours, the materials and the sizes. Our results demonstrate that a useful category hierarchy can emerge based on the constructional behaviour captured by the categorial network of an agent.

5 Related Work

Many computational models for grounded language acquisition have been developed in different fields, including cognitive linguistics, AI and robotics. This paper studies this grand challenge from a constructionist perspective. Therefore, in what follows, we outline the different strands of work that take this perspective. Following a recent survey by Doumen et al. (2025), we thus focus on constructionist models that incrementally acquire productive form-meaning mappings that extend beyond lexical items. Doumen et al. (2025) distinguish approaches based on how much semantic supervision is provided. In a first set of models, training examples pair an utterance with its gold semantic annotation (e.g. Alishahi and Stevenson, 2008; Beuls et al., 2010; Chang, 2008; Doumen et al., 2024; Gerasymova and Spranger, 2010, 2012). Other models reduce this supervision by presenting multiple candidate gold semantic annotations, introducing referential uncertainty (Abend et al., 2017; Beekhuizen and Bod, 2014; Beekhuizen, 2015; Chen and Mooney, 2008; Dominey, 2005a,b; Dominey and Boucher, 2005; Garcia Casademont and Steels, 2015, 2016; Gaspers et al., 2011; Gaspers and Cimiano, 2012, 2014; Gaspers et al., 2017; Kwiatkowski et al., 2010, 2011, 2012; Steels, 2004). A third set of models focus on learning in situated interactions without gold semantic annotations altogether. In these works, a combination of a predefined lexicon, categories or ontology is assumed (Artzi and Zettlemoyer, 2013; Nevens et al., 2022; Spranger, 2015; Spranger and Steels, 2015; Spranger, 2017). Finally, De Vos et al. (2024) present a grammar coupled with concepts grounded in a way similar to Section 2.1. Notably, their approach is applied to the same visual question answering task considered in this paper. However, whereas they manually designed a grammar tailored to the task, our focus lies on the acquisition of a grammar across different levels of abstraction. This makes the problem significantly more challenging and motivated our use of a subset of the dataset (see Section 4). As such, direct comparison is not straightforward. De Vos

et al. (2024) report an accuracy of 96% on the full dataset, while we achieve near-perfect success on the subset.

A growing related strand of research examines to what extent large language models (LLMs) capture constructional knowledge. These probing studies indicate that state-of-the-art LLMs can capture substantive constructions reasonably well, but have more difficulty with more schematic patterns (see e.g. Weissweiler et al. (2022); Bonial and Tayyar Madabushi (2024); Zhou et al. (2024); Rozner et al. (2025)). These findings provide valuable insights into the strengths and limitations of current models. Our objective, rather, is to present a mechanistic model in which constructions at varying levels of schematicity emerge incrementally through situated communicative interactions, rather than via optimisation for next-token prediction over large-scale corpora.

6 Discussion and Conclusion

This paper has presented a computational model in which a construction grammar is acquired directly from sensory experience, capturing constructions at varying levels of abstraction. We have integrated two previously separate but related experiments operationalised in the language game paradigm, guided by the hypothesis that constructions at different levels can be acquired through the same underlying cognitive mechanisms. While Beuls and Van Eecke (2024) formulated this idea at a conceptual level, we offer a concrete operationalisation. In our approach, constructions are acquired through repeated situated communicative interactions between a tutor and a learner agent. Across these interactions, the learner identifies regularities (whether these are associations between sensory feature values and linguistic forms or correspondences between syntactic patterns and semantic operations) and uses those regularities to incrementally refine its linguistic system. To enable this integration, we have introduced a component that induces a category hierarchy from the slot-filler relations of the acquired constructions, thereby replacing a major scaffold of the earlier model by Nevens et al. (2022), which assumed access to a predefined hierarchy. In this setting, the component derives category hierarchies that are one layer deep, although future work could investigate extensions to multi-level hierarchies.

The methodology has been validated through an experiment in the synthetic CLEVR environment. The experiment has demonstrated that lexical constructions that were grounded in the sensors of the agent and were acquired in referential games can serve as building blocks for abstract grammatical constructions in a subsequent question-answering game. In this paper, we focused on the acquisition of lexical and item-based constructions. These results provide computational support for a core assumption in construction grammar, showing how both purely substantive and more abstract constructions can emerge from repeated situated communicative interactions. However, further work is needed to investigate the acquisition of constructions at all levels of abstraction in more complex environments.

Acknowledgements

We would like to thank Liesbet De Vos, Jamie Wright, Arno Temmerman and the anonymous reviewers for their valuable comments on earlier versions of the paper. The research reported on in this paper was funded by the F.R.S.-FNRS-FWO WEAVE project HERMES I under grant numbers T002724F (F.R.S.-FNRS) and G0AGU24N (FWO), the Flemish Government under the Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen programme and the AI Flagship project ARIAC by DigitalWallonia4.ai.

References

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

Afra Alishahi and Suzanne Stevenson. 2008. A computational model of early argument structure acquisition. *Cognitive Science*, 32(5):789–834.

Yoav Artzi and Luke Zettlemoyer. 2013. Weakly supervised learning of semantic parsers for mapping instructions to actions. *Transactions of the Association for Computational Linguistics*, 1:49–62.

Barend Beekhuizen. 2015. *Constructions Emerging*. Ph.D. thesis, Universiteit Leiden.

Barend Beekhuizen and Rens Bod. 2014. Automating construction work: Data-oriented parsing and onstructivist accounts of language acquisition. In Ronny Boogaart, Timothy Colleman, and Gijsbert Rutten, editors, *Extending the Scope of Construction Grammar*, pages 47–74. De Gruyter Mouton, Berlin, Germany.

Heike Behrens. 2009. Usage-based and emergentist approaches to language acquisition. *Linguistics*, 47(2):383–411.

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 and Sebastian Höfer. 2011. Simulating the emergence of grammatical agreement in multi-agent language games. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pages 61–66, Washington, D.C., USA. AAAI Press.

Katrien Beuls and Paul Van Eecke. 2024. Humans learn language from situated communicative interactions. What about machines? *Computational Linguistics*, 50(4):1277–1311.

Katrien Beuls and Paul Van Eecke. 2025. Construction grammar and artificial intelligence. In Mirjam Fried and Kiki Nikiforidou, editors, *The Cambridge Handbook of Construction Grammar*, pages 543–571. Cambridge University Press, Cambridge, United Kingdom.

Claire Bonial and Harish Tayyar Madabushi. 2024. Constructing understanding: on the constructional information encoded in large language models. *Language Resources and Evaluation*.

Jérôme Botoko Ekila, Jens Nevens, Lara Verheyen, Katrien Beuls, and Paul Van Eecke. 2024. Decentralised emergence of robust and adaptive linguistic conventions in populations of autonomous agents grounded in continuous worlds. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multi-Agent Systems*, pages 2168–2170, Richland, SC, USA. IFAAMAS.

Joan Bybee. 2010. *Language, usage and cognition*. Cambridge University Press, Cambridge, United Kingdom.

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

David L. Chen and Raymond J. Mooney. 2008. Learning to sportscast: a test of grounded language acquisition. In *Proceedings of the 25th International Conference on Machine learning*, pages 128–135.

William Croft. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press, Oxford, United Kingdom.

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, Singapore. World Scientific.

Liesbet De Vos, Jens Nevens, Paul Van Eecke, and Katrien Beuls. 2024. Construction grammar and procedural semantics for human-interpretable grounded language processing. *Linguistics Vanguard*, 10(2):565–574.

Holger Diessel. 2017. Usage-based linguistics. In Mark Aronoff, editor, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, Oxford, United Kingdom.

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

Jonas Doumen, Katrien Beuls, and Paul Van Eecke. 2024. Modelling constructivist language acquisition through syntactico-semantic pattern finding. *Royal Society Open Science*, 11(7):231998.

Jonas Doumen, Veronica J. Schmalz, Katrien Beuls, and Paul Van Eecke. 2025. The computational learning of construction grammars: State of the art and prospective roadmap. *Constructions and Frames*, 17(1):141–174.

Ewa Dąbrowska and Elena Lieven. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, 16(3):437–474.

Charles J. Fillmore. 1988. The mechanisms of "construction grammar". In *Annual Meeting of the Berkeley Linguistics Society*, volume 14, pages 35–55.

Emília Garcia Casademont and Luc Steels. 2015. Usage-based grammar learning as insight problem solving. In *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science*, pages 258–263.

Emília Garcia Casademont and Luc Steels. 2016. Insight grammar learning. *Journal of Cognitive Science*, 17(1):27–62.

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

Judith Gaspers and Philipp Cimiano. 2014. A computational model for the item-based induction of construction networks. *Cognitive Science*, 38(3):439–488.

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

Judith Gaspers, Philipp Cimiano, Katharina Rohlfing, and Britta Wrede. 2017. 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.

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. IOS Press.

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, NY, USA.

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

Adele E. Goldberg. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5):219–224.

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 *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2901–2910.

Frédéric Kaplan, Luc Steels, and Angus McIntyre. 1998. An architecture for evolving robust shared communication systems in noisy environments. Technical report, Sony CSL Paris.

Tom Kwiatkowski, Sharon Goldwater, Luke Zettlemoyer, and Mark Steedman. 2012. A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their meanings. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 234–244. Association for Computational Linguistics.

Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2010. Inducing probabilistic CCG grammars from logical form with higher-order unification. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1223–1233. Association for Computational Linguistics.

Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2011. Lexical generalization in CCG grammar induction for semantic parsing. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1512–1523. Association for Computational Linguistics.

George Lakoff. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.

Ronald W. Langacker. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*, volume 1. Stanford University Press, Stanford, CA, USA.

Elena Lieven. 2014. First language learning from a usage-based approach. In Thomas Herbst, Hans-Jörg Schmid, and Susen Faulhaber, editors, *Constructions Collocations Patterns*, pages 9–32. De Gruyter Mouton, Berlin, Germany.

Martin Loetzsch. 2015. *Lexicon formation in autonomous robots*. Ph.D. thesis, Humboldt-Universität zu Berlin, Berlin, Germany.

Jens Nevens. 2022. *Representing and learning linguistic structures on the conceptual, morphosyntactic, and semantic level*. Ph.D. thesis, Vrije Universiteit Brussel, Brussels: VUB Press.

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.

Jens Nevens, Paul Van Eecke, and Katrien Beuls. 2020. From continuous observations to symbolic concepts: A discrimination-based strategy for grounded concept learning. *Frontiers in Robotics and AI*, 7(84).

Eleanor Rosch, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439.

Joshua Rozner, Leonie Weissweiler, Kyle Mahowald, and Cory Shain. 2025. Constructions are revealed in word distributions. *arXiv preprint arXiv:2503.06048*.

Gerard Salton and Michael J. McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw Hill.

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. IEEE.

Michael Spranger. 2017. Usage-based grounded construction learning: A computational model. In *The 2017 AAAI Spring Symposium Series*, pages 245–250, Washington, D.C., USA. 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, Washington, D.C., USA. AAAI Press.

Luc Steels. 1995. A self-organizing spatial vocabulary. *Artificial Life*, 2(3):319–332.

Luc Steels. 1999. *The Talking Heads experiment: Volume I. Words and Meanings*. Best of Publishing, Brussels, Belgium.

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. Association for Computational Linguistics.

Luc Steels and Emília Garcia Casademont. 2015. Ambiguity and the origins of syntax. *The Linguistic Review*, 32(1):37–60.

Luc Steels, Paul Van Eecke, and Katrien Beuls. 2022. Usage-based learning of grammatical categories. *arXiv preprint arXiv:2204.10201*.

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

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.

Remi van Trijp. 2008. The emergence of semantic roles in fluid construction grammar. In *The Evolution of Language*, pages 346–353. World Scientific.

Leonie Weissweiler, Valentin Hofmann, Abdullatif Köksal, and Hinrich Schütze. 2022. The better your syntax, the better your semantics? Probing pretrained language models for the English comparative correlative. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10859–10882. Association for Computational Linguistics.

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

William A. Woods. 1968. Procedural semantics for a question-answering machine. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, pages 457–471, New York, NY, USA.

Shijia Zhou, Leonie Weissweiler, Taiqi He, Hinrich Schütze, David R. Mortensen, and Lori Levin. 2024. Constructions are so difficult that even large language models get them right for the wrong reasons. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3804–3811. Association for Computational Linguistics.