Usage-based Grammar Learning as Insight Problem Solving

Emilia Garcia-Casademont[2]
Luc Steels[1,2]
[1] ICREA [2] Institut de Biologia Evolutiva (UPF-CSIC)
Dr. Aiguader 88, Barcelona 08003, Spain

Abstract

We report on computational experiments in which a learning agent incrementally acquires grammar from a tutoring agent through situated embodied interactions. The learner is able to detect impasses in routine language processing, such as missing a grammatical construction to integrate a word in the rest of the sentence structure, to move to a meta-level to repair these impasses, primarily based on semantics, and to then expand or restructure his grammar using insights gained from repairs. The paper proposes a cognitive architecture able to support this kind of insight learning and tests it on a grammar learning task.

Keywords: Usage-based language learning; insight problem solving; Fluid Construction Grammar.

What is usage-based language learning

Many researchers have argued for a usage-based approach (Langacker, 2000) to language learning based on empirical data from child language acquisition (Tomasello, 1992). This approach emphasizes that learning is essentially data-driven and gradual (Bybee, 2006). It takes place in embodied situated interactions in which not only the utterance and its syntactic properties (e.gthe ordering of the words) but also the underlying meanings and communicative context are available to the learner.

The usage-based approach goes hand in hand with a constructional perspective on grammar (Goldberg, 2006), which emphasizes that language structure is motivated by usage, instead of innate, arbitrary, formal principles of universal grammar. Grammar learners have to discover what role the grammatical markers and syntactic patterns play in expressing meaning and achieving communicative functions, and how syntax helps to dampen combinatorial explosions and avoid ambiguity. Construction grammar therefore argues that the fundamental unit of grammar is the construction, a sign that relates meaning or function with form with the intermediary of syntactic and semantic categorizations. A construction is very rich, packaging constraints at different layers of language (from phonetics and phonology to morphology, syntax, semantics and pragmatics) and from many different perspectives (phrase structure, functional structure, argument structure, temporal structure, information structure, etc.). Constructions are acquired in a gradual data-driven way with learners creating progressively a huge inventory of more sophisticated constructions linked in networks (Tomasello, 1992).

Computational simulations of usage-based constructional learning are rare, despite widespread support for a usage-based approach in cognitive science circles and a growing success of construction grammar in empirical linguistics, studies of child language, historical linguistics and other areas of language studies - even though there is some exceptional work, such as by Nancy Chang (Chang, 2008). This computational gap is undoubtly due to the fact that the field of computational construction grammar is still in its infancy and not enough computational research has been done so far on possible learning strategies.

Nevertheless computer simulations are a valuable complement to the empirical work on language learning. They would help us a great deal to get a much clearer idea of what constructions look like from a computational point of view and how they are acquired. Computer simulations allow us to test the role of different cognitive mechanisms for grammar learning by performing experiments with and without these mechanisms and by varying the learning challenges and the nature and amount of the input the learner gets.

Insight Problem Solving

Today, most machine-learning of language uses a form of Bayesian unsupervised grammar learning operating over large amounts of data (Bod & Scha, 2003). Because these models are data-driven, they subscribe to one of the key tenets of a usage-based approach to grammar learning and they have therefore already been used for simulating child grammar acquisition (Bannard, Lieven, & Tomasello, 2009). But although this approach is obviously relevant, we explore here a complementary learning method which views grammar learning as a form of insight problem solving.

Insight problem solving is familiar from the psychological literature on problem solving (Ohlsson, 1984). It makes a distinction between two modes: routine and meta level problem solving. For routine problem solving, the problem solver uses a set of micro-operations that either provide a direct solution or are a step towards a solution. The main challenge for the problem solver is to find a path between the initial problem state and the goal in a search space of possible hypotheses.

Meta-level problem solving is necessary when the problem solver reaches an impasse. The given statement of the problem (as understood by the problem solver), the representations being used, and the known operators do not allow a straightforward solution. Meta-level operators may then unblock the impasse. For example, they may reinterpret the problem statement by relaxing some of its constraints (as needed for solving the 9-dot problem), they may change internal representations inferred from the problem (as in the Anthony and Cleopatra problem (Patrick & Ahmed, 2014)), or possibly introduce new operators. One very common meta-operator, studied in particular by Koehler in his classic insight learning with chimpanzees, is to coerce objects to have functions that they normally do not have (Koehler, 1956). For example, to view a shoe as a hammer so that it can play the role of the instrument in a hitting action.

In the case of language, the micro-operations for routine problem solving constitute the application of constructions to expand a transient structure (which captures a particular hypothesis for comprehending or producing an utterance) to a more extended transient structure. For example, a construction might combine an article and a noun into a noun phrase. Although some exploration and backtracking may be necessary to know which construction needs to be applied, the problem solver is in principle able to solve the problem, i.e. to reconstruct the meaning of an utterance as listener or to produce an utterance expressing the target meaning as speaker.

An impasse here means, for example, that the hearer encounters an unknown word, a particular word does not fit within the syntactic pattern implied by its immediate context, an ordering constraint is violated, no structure can be found that integrates all words in the utterance, or there is syntactic and semantic ambiguity implying that some grammatical signals preventing ambiguity have not been picked up properly. These impasses are frequent when learning a new language. The speaker may also reach an impasse because, for example, he may lack a word to express a particular meaning, a word he would like to use does not fit with a construction already chosen, the available constructions may leave significant ambiguity, etc.

Meta problem solving for language involves a set of meta-operators and a cognitive architecture as in Figure 1. For example, the listener could try to handle a sentence which he cannot parse by ignoring possible errors in it (e.g. the wrong word order). Or he may handle the presence of an unknown word by inferring from the syntactic context, the situation model and the ontology, what a possible meaning might be. The speaker could coerce a word to have an additional lexical category so that it fits, as in “He googled him” where the noun ”google” has been coerced into a verb.

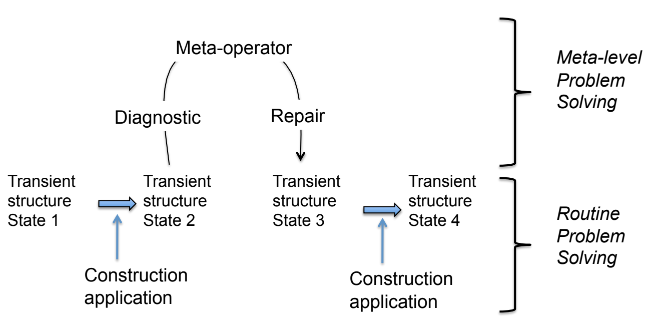

Figure 1: Cognitive architecture supporting insight language learning. There is a layer of routine problem solving, diagnostics continuously monitoring activity at this layer in order to discover an impasse, and meta-level operators try to repair the impasse, for example by coercing a word to fit with an available grammatical construction.

But that is not the end of the story. After solving a problem using meta-operators, a learner should then try to expand his available repertoire of routine operators (i.e. constructions) to capture the insight gained from dealing with the impasse to ensure that in the future the impasse does not occur anymore. When that is the case, we talk about insight learning (Koehler, 1956). It is based on a set of learning-operators that make small changes to the grammar, for example, add a new lexical category to the lexical construction of a word or make a construction more general by relaxing some constraints on one of its slots.

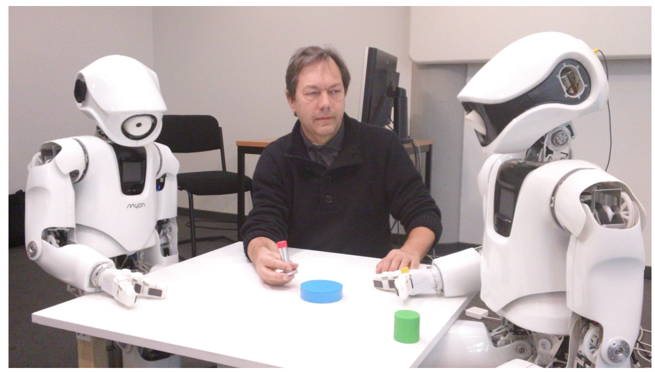

Figure 2: Example of interaction between two agents, one acting as tutor and the other as learner. A human experimenter creates scenes about which the agents can converse. The robots are equiped with sophisticated components for perception, action, world modeling, and script following in order to play a language game. Grammar learning and tutoring takes place within the context of such games.

The rest of the paper reports on experiments attempting to computationally implement insight language learning. Our goal is not to handle the full complexity of human language, that would not be possible at this point and the results would be very hard to analyse, but rather, to **create a minimal model** in order to test the feasibility of the approach and examine the first concrete examples of meta-operators and learning-operators. We therefore use a minimal representation of meaning and focus on a minimal form of syntactic structure, namely phrase structure.

The experiments are intended for the MYON robotic platform shown in Figure 2 (Steels & Hild, 2012). There are two humanoid robots, one acting as tutor and the other one as learner. The situation is manipulated by an experimenter which puts objects on the table (e.gà small piece of paper, a wooden block, etc.) and performs actions with them. The robots are able to perceive the situation before them, segment the objects and detect relations, such as spatial relations or movements. They can also point to objects and gesture success or failure in the language game. The relations are compositional (as in ”a black round wooden block”) with unlimited recursion (as in ”a small paper which is on a small paper which is on the table”). The situation model can also include mental states (e.g. ”Jorge wants the wooden block (to be) on the table”).

The next section first describes routine language processing. Then we turn to meta-level problem solving, discussing syntactic and semantic expansions, and to learning operators. The paper concludes with some experimental results. There is an on-line web demo and additional material available at www.biologiaevolutiva.org/lsteels/demos/EAP-garcia-steels-2015/

'Routine' Language Processing

Meaning representation

The complex visual processing and conceptualization processes required for these experiments have been described at length in (Spranger, Loetzsch, & Steels, 2012). The situation models of speaker and hearer are not identical because their perception is different, but in general they are sufficiently shared that communication is possible, otherwise the language game fails. For the purposes of the experiments reported here, we have defined a 'situation model generator' that generates possible situation models in the same format as obtained by visual processing.

For the representation of the situation model we use well-known standard techniques from (typed second order) logic. The situation model is internally represented in terms of n-ary predicates, which have objects perceived in the world as arguments. The objects are labeled *obj-1*, *obj-2*, etc. The objects can be bound to variables which are written down with a question mark in front, as in *?obj-1*, *?obj-2*, etc. Each predicate consists of an attribute, that acts also as a type, and a value.

Unary predicates are written down as:

(attribute value object-to-which-predicate-applies)

as in

(color red obj-1) or *(material plastic ?obj-6)*

In the first example, the color *red* is predicated about *obj-1*, i.e. obj-1 is red in the current situation model. In the second example, the *material* property *plastic* is predicated about a variable object *?obj-6*. It will be true if there is an object in the world which has the material property plastic.

N-ary predicates are decomposed into a number of separate predicates. One for the predicate itself and then predicates for each of the arguments. For example, suppose there is a predicate of type *moving* (which is for all types of movement) with a possible value *away*, then there are two predicates for its arguments, as illustrated in the following example.

(moving away ?r-2) ; the main predicate
(away-arg-1 ?r-2 ?o-3) ; the object that is moving
(away-arg-2 ?r-2 ?o-1) ; the object being moved away from.
?r-2 gets bound to the event of moving, *?o-3* to the object that is moving and *?o-1* to the object *?o-3* moves away from.

Different predications can be combined into conjunctions and they are linked together by reusing the same variable-names or object-names. For example, the utterance "a small paper moves away from a wooden table" would be represented as

(moving away ?r-2) ; the main predicate
(away-arg-1 ?r-2 ?o-3) ; the object that is moving
(away-arg-2 ?r-2 ?o-1) ; away from
(size small ?o-3) ; the moving object is small
(material paper ?o-3) ; and its material is paper
(physobj table ?o-1) ; the movement is away from a table
(material wood ?o-1) ; which is made of wood

The different linkages between the predications through their arguments equalities may be represented graphically as a semantic network (See Figure 3). One of the objects in this network is the topic of the utterance, for example, the event itself (i.e. *?r-2*), or the object which is moving (i.e. *?o-3*) as in the utterance "the small paper moving-away from the wooden table".

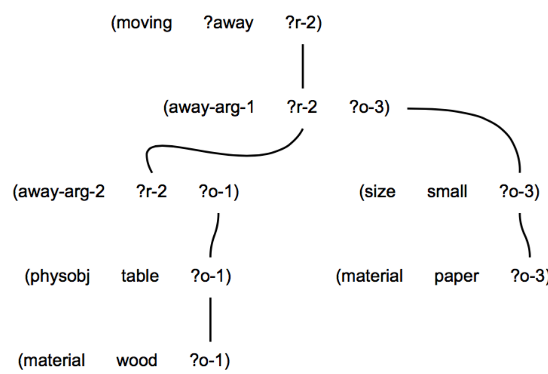

Figure 3: Semantic network representing the meaning of an utterance. Each node is a predication and the links represent co-referential relations between the arguments.

To simplify the experiments, we use utterances with only content-carrying lexical items and only word order and phrase structure as the means for expressing syntactic structure, ignoring other syntactic devices like morphology, grammatical function words, agreement, etc. Thus, the utterance "a small paper moves away from a wooden table" would be rendered as "small paper moves-away-from wooden table". There is of course no necessity to use English-like words, that is only done to make the utterances understandable for the reader, and there is no reason why the grammar will be English-like, except that English also makes heavy use of phrase structure.

Grammar representation

We use Fluid Construction Grammar (FCG) for the representation of the grammar (Steels, 2011). FCG views language processing in terms of operations over transient structures. A transient structure captures all that is known about a particular utterance being parsed or produced. In routine language processing, transient structures are expanded by the application of constructions in a process of matching (to see whether the construction can apply) and merging (to add information from the construction to the transient structure).

FCG represents transient structures in terms of feature structures, similar to many other computational formalisms in use today, such as HPSG. A feature structure consists of a set of units which correspond to words and phrases, and features associated with these units. The features can be at any level of linguistic description. For the present experiments, features of a unit include: *meaning* (the set of predications), *referent* (the object referred to by the unit), *form* (the strings and ordering relations), *args* (the arguments of these predications which can be used to combine this unit with the meaning of other units), *sem-cat* (the semantic categorizations), *boundaries* (the left and right boundaries), *sem-subunits* and *syn-subunits* (the constituents), *potential-syn-cat* (the potential lexical categories (parts of speech)), *syn-cat* (the actual lexical category or the phrasal category), *head* (the subunit acting as the head of the phrase), *footprints* (constructions that changed this unit).

A construction is an association between meaning and function (the semantic pole) and form constraints (the syntactic pole). Lexical constructions associate one or more predicates and a semantic categorization of the predicates (equal to the attribute (i.e. type) of the predicate) in the semantic pole with the occurrence of a word-string and lexical categories (i.e. parts of speech) of that word in the syntactic pole. Grammatical constructions, in this case restricted to phrase structure constructions, associate a pattern of units with semantic constraints in the semantic pole with syntactic categories (lexical or phrasal categories) and word orders in the syntactic pole. Each construction has a score (between 0.0 and 1.0) which reflects the success of that construction in past language interactions.

Constructions in FCG consist of two parts. A conditional part (written on the right hand side) which specifies under which conditions a construction can become active and a contributing part (written on the left hand side) which specifies what the construction contributes to the transient structure. Constructions must be usable both in comprehension and production. So the conditional part is split into two 'locks'. A production lock (on top) which has to match with the transient structure in production and a comprehension lock (below it) which has to match in comprehension. When a lock fits with a transient structure all the information of the construction (the other lock and the contributing part) gets merged in. The fact that constructions in FCG can be used both for comprehension and production is crucial for insight learning because once the learner has been able to deduce the meaning (possibly indirectly), he can produce the same meaning himself with his own inventory and thus compare it to the utterance produced by the speaker.

A lexical construction for the word "paper" is:

$$\left[\begin{array}{l}\textbf{?word}\\ \hline\hline \text{sem-cat: material}\\ \text{potential-syn-cat:}\\ \quad\{\text{adj, noun}\}\\ \text{referent: ?obj}\end{array}\right] \leftarrow \left[\begin{array}{l}\textbf{?word}\\ \hline\hline \text{\# meaning:}\\ \quad\{(\text{material paper ?obj})\}\\ \hline \text{\# form:}\\ \quad\{(\text{string ?word paper})\}\end{array}\right]$$

"Paper" has two potential lexical categories: adjective (as in "a paper basket") and noun (as in "a small paper"). Here is a simplified example of a grammatical construction:

$$\left[\begin{array}{l}\textbf{?np-unit}\\ \hline\hline \text{constituents:}\\ \quad\{\text{?word-unit-1, ?word-unit-2}\}\\ \text{sem-cat: material}\\ \text{syn-cat: noun-phrase}\\ \text{head: word-unit-2}\\ \text{referent: ?obj}\end{array}\right] \leftarrow$$

$$\left[\begin{array}{l}\textbf{?word-unit-1}\\ \hline\hline \text{sem-cat: size}\\ \text{referent: ?obj}\\ \hline \text{potential-syn-cat:}\\ \quad\{\text{adjective}\}\end{array}\right] \leq \left[\begin{array}{l}\textbf{?word-unit-2}\\ \hline\hline \text{sem-cat: material}\\ \text{referent: ?obj}\\ \hline \text{potential-syn-cat:}\\ \quad\{\text{noun}\}\end{array}\right]$$

Meta-level processing

The consecutive application of constructions expands the transient structure to go from meaning to form or vice versa. But occasions will arise when no construction can be applied, particularly in language learning. The listener then moves to the meta-level to repair the situation and then consolidate the outcome of the repair.

Syntactic meta-operators

The listener can try to find a construction which is only partially matching, and either coerce words to fit into that construction (syntactic coercion) or expand the applicability of the construction (extension). More specifically,

+ **Coercion:** A construction is found that is semantically compatible but one or more words do not have the appropriate lexical category (as in the example of "googled" where a noun occurs in a context where a verb is expected). The meta-operator then adds the lexical category to the word-unit in the transient structure and the construction then applies.

+ **Extension:** A construction is found for which all the components match but there is another word ordering. In this case, the ordering constraint can be relaxed and the construction applied.

+ **Reduction:** All components of an existing construction could be matching with the transient structure but there is a superfluous unit and no matching construction without this unit. This superfluous unit can be ignored and the construction as well as possible applied.

Semantic meta-operators

When no partially matching constructions could be found, it may be possible to use the situation model and combine units for words or phrases based on semantics, specifically:

+ **Build-or-extend-group**: If two words or word groups expressing unary predicates refer to the same object, they can be combined. For example, if there is a group for *wooden table* (based on an existing construction) and the utterance is *small wooden table*, the word-unit for *small* can be linked in with the group-unit for *wooden table*. The group-unit retains the same referent.

+ **Build-hierarchy** When a relational word is encountered, i.e. a word which introduces a predicate with more than one argument, such as *moves-away-from*, and no constructions are available to integrate it in the transient structure, the meta-operator looks in the world-model to detect which object plays which role and then combines the units for these objects into a new hierarchical unit. The meta-operator also needs to decide which of the arguments is going to be the referent. In principle, it could be any argument (e.g. the event of moving, the mover, or the object being moved away from). In practice, the referent is determined by the semantic network that is expressed. For example, in the sentence ”the ball on the table”, the referent of the unit based on the relational word ”on” is the ball, whereas in ”He wants the ball (to be) on the table” the referent is the on-relation itself.

Consolidation

When an utterance could be successfully parsed after one or more repairs, the learner activates consolidation operators that integrate the insights that were obtained into his construction inventory. In some cases this is straightforward, for example, coercion can be consolidated by adding the relevant lexical category to the potential lexical categories of a word. In other cases, more complex manipulations are required. Existing constructions are first copied and then additions and changes made.

Alignment

The meta-operators and learning-operators are hypotheses made by the learner about the language of the speaker and mistakes are unavoidable. The learner therefore needs an additional mechanism to progressively discard wrong hypotheses based on further input. We have achieved this by updating the score of constructions using the well known lateral inhibition learning rule. Knowing which constructions c_i need an increased score is easy: they are the constructions that were used on the path towards the final transient structure. We use the following update rule: $\sigma_{c_i} \leftarrow \sigma_{c_i}(1-\gamma)+\gamma$, with $\gamma = 0.1$ a constant. What competing constructions c_j need to be decreased? First of all, all constructions that started off a wrong branch in the search space during comprehension, i.e. a branch which is not on the path to the final solution. Next, the listener produces himself the utterance based on the meaning deduced from the comprehension process and then finds all constructions that start off a wrong branch while producing. Their scores need to be decreased as well. We use the following update rule: $\sigma_{c_j} \leftarrow \sigma_{c_j}(1-\gamma)$.

Results

We now report on two (of many more) experiments exercising the cognitive architecture in Figure 1 and the meta- and learning-operators from the previous section. We have implemented a tutor which is initialized with a lexicon of 40 lexical constructions and a grammar with 30 grammatical constructions. The tutor grammar includes adverbs, adjectives, nouns, verbs, prepositions, pronouns and relative pronouns as well as noun phrases of different levels of complexity, verb phrases, main clauses and relative clauses. The tutor produces a stream of utterances each describing a particular topic (object or event) in a scene. Some example utterances are “”Paul sees (the) red carpet (that) Emilia wants”, or “Paul believes (that) Emilia wants (the) carpet on (the) big wooden table.”. The learner is initialized with the same lexicon (because we focus here on the acquisition of grammar). He is endowed with the various operators described above, but without any grammatical constructions or syntactic categories (parts of speech and phrases).

In a concrete interaction (following the classical Naming Game), tutor and learner share the same situation model (represented as a semantic network as in Figure 2). The tutor then chooses a topic (one object to be referred to) and produces a description of this topic using his lexicon and grammar. Then the learner tries to parse the utterance and interpret the extracted meaning against the situation model. The interaction is successful when the learner was able to identify the topic. This may involve both routine and meta-level processing followed by learning. Each experiment is carried out for 5 different tutor-learner pairs, using random choices from a set 20 situations, so that results can be compared.

Experiment 1. Lexical categories given

The first experiment tests out the learning operators assuming that the learner already knows the lexical categories, i.e. parts of speech, of the words in the lexicon. Moreover grammatical constructions are limited to those having only one semantic category and one syntactic category per subunit. The task is to learn the grammatical constructions. Figure 4 shows the results, measuring communicative success, inventory size (the number of constructions of the learner) and alignment (how often the listener’s reproduction of the meaning of an utterance is equal to the utterance of the speaker).

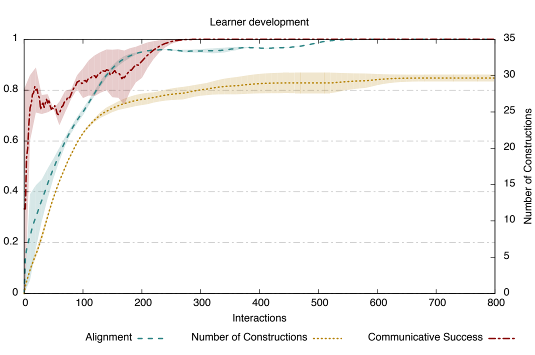

Figure 4: Grammar learning with lexical categories known. 800 language games are shown (x-axis). The learner reaches the same number of grammatical constructions (namely 30 on right y-axis) and a total alignment (left y-axis), demonstrating that he successfully acquired the grammar. The shading around the lines represent the confidence interval of 5 runs.

Experiment 2. Multiple semantic categories

The second experiment assumes that the learner does not know any a priori lexical categories. This obviously makes the learning problem harder. Also, grammatical constructions can have more than one semantic category per slot, which means that constructions can be more general. They still have only one syntactic category for slots, whereas individual words can have several lexical categories (syncretism). Because constructions can be more general, we end up with a smaller inventory.

Results for inventory size and alignment are shown in Figure 5. The graph also shows syntactic ambiguity (the number of superfluous branches when applying constructions plus the number of possible combinations in the semantic meta-operators divided by the number of words in the utterance), and semantic ambiguity (the number of situations considered but cut off in erroneous branches divided by the number of variables introduced by all words). Thanks to grammar both types of ambiguity get drastically reduced, which proves that an important function of grammar is to dampen combinatorial search.

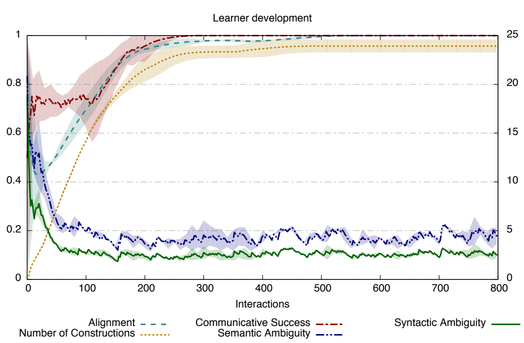

Figure 5: Grammar learning without lexical categories known. The learner reaches fewer grammatical constructions (25) and a total alignment, although this takes longer compared to Figure 4.

Conclusion

The paper explored the hypothesis that usage-based grammar learning can be modeled as an insight problem solving process. The computational simulations showed that this is possible requiring (i) a strong semantic component to make meaning-driven learning possible, (ii) a meta-level architecture similar to SOAR or ACT-R and (iii) appropriate learning and alignment operators.

Of course, the model is much simpler compared to the full complexity of the resources that human language learners bring to bear on language acquisition. However, the advantage of a simpler system is that the minimal characteristics of a language learner become more starkly visible. Moreover this method allows us to causally examine the impact of operators, and thus supports empirical research into which operators are available to human language learners. The simulations have already been scaled up to sentences with much higher complexity, including full recursion, and will be tested against corpora of child-directed speech, such as CHILDES, in the future.

Acknowledgments

The research reported here has been funded by an ICREA Research fellowship to LS and a Marie Curie Integration Grant EVOLAN. The project has received further funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 308943, which is the FET-OPEN Insight project. We are indebted to Paul Van Eecke for comments.

References

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *PNAS*, *106*(41), 17284-9.

Bod, R., & Scha, R. (2003). *Data-oriented parsing*. Stanford, Palo Alto: Center for the Study of Language and Information.

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*. Berkeley: Computer Science Division, University of Berkeley.

Goldberg, A. E. (2006). *Constructions at work, the nature of generalization in language*. Great Clarendon Street, Oxford: Oxford University Press.

Koehler, W. (1956). *The mentality of apes*. New York: Routledge and Kegan Paul.

Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1–63).

Ohlsson, S. (1984). Restructuring revisited: Ii. an information processing theory of restructuring and insight. *Scandinavian Journal of Psychology*, *25*, 117-129.

Patrick, J., & Ahmed, A. (2014). Facilitating representation change in insight problems through training. *Journal of Experimental Psychology. Learning Memory and Cognition*, *40*(2), 532-543.

Spranger, M., Loetzsch, M., & Steels, L. (2012). A Perceptual System for Language Game Experiments. In L. Steels & M. Hild (Eds.), *Language Grounding in Robots* (pp. 89–110). Springer.

Steels, L. (Ed.). (2011). *Design patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Steels, L., & Hild, M. (Eds.). (2012). *Language grounding in robots*. New York: Springer.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.