Towards a Computational Model for Early Language Acquisition

Andreas van Cranenburgh, Arjan Nusselder, Nadya Peek and Carsten van Weelden
{acranenb, anussel, npeek, cweelden}@science.uva.nl

June 30, 2007

Abstract

How do humans learn language? According to the usage based model of construction grammar, language is inductively learned. Linguistic input can be interpreted by making analogies with previously seen language. However, how does one start to interpret linguistic input before one has experiences to compare it to? In this paper, a computational model for building this first corpus using construction grammar is introduced. An implementation for simulation is presented, along with some promising experimental results.

1. Introduction

Human language differs from animal communication in several ways. Language is symbolic– it can refer to anything in the world, observable or unobservable, thereby allowing hypothetical situations. Language is grammatical– it assigns individual meanings to words, but also new meanings to particular series of words. Language is acquired during individual development– its form depends greatly on the individual's surroundings, and is certainly not uniform for everyone in the world [10].

It is unknown when language originated, and why it is not observable yet in other species, or why it is so inherently local. Its origin is a classic continuity paradox: the human way of communicating with symbols may have been a trigger for the evolution of language, or the onset of language may have enabled symbolic communication.

Either way, each individual's acquisition of language portrays its own miniature evolution. Every human learns the linguistic behavior of the community it is raised in. Therefore, every human needs to be flexible enough at birth to be able to deal with any of the variants of language that exist on earth [10].

A language begins with words. A child might learn the meaning of a word by hearing the word while observing its semantic context in reality, and thus be able to map a word to a particular situation. Unfortunately this immediately runs us into a frame problem. Even if the child manages to discern only one word, how will it know which part of reality the word would map to?

Even while children are still trying to extract meaning from adult utterances, they can already start learning syntactical structure. Learning the first syntactic constructions does not differ much from learning the first words– a child is left to observe events while trying to discern the speaker's communicative intention. Every language has its own grammatical conventions which all need to be learnable.

One could even argue that children do not necessarily segment language exactly at word boundaries, and that the smallest constituents of language are not single words or morphemes, but small grammatical constructions [2].

Regardless, as the child is learning its first words and grammatical constructions, it makes many mistakes. How are these corrected? Does it remember all of the situations it has seen before? How does it learn word order? In short, how is language represented in the mind of a child?

2. Background

There is no consensus on how language is stored in the mind. There have been neurobiological advances in understanding where language is processed and produced in the brain [7], but there is still disagreement on *how* the language is processed and produced. The largest players in the discussion are the generative grammarians (Chomsky, Pinker) and the construction grammarians (Kay, Chang, Fillmore, Tomasello). In this section we will explain how both camps see the storage and acquisition of language within their framework.

2.1. Generative grammar

Generative grammar is partially inspired by formal grammar, and describes the syntactic structure of natural language. It was introduced by Noam Chomsky in the 1950s [5]. It has an autonomous syntax module which generates well-formed word sequences by means of rewrite rules which are applied to abstract syntactic categories, such as *Verb Phrases* or *Noun Phrases*. This syntax producing module is said to contain the same rules in all humans, and these rules are known as the Universal Grammar.

According to Chomsky, the Universal Grammar is innate, which is why all languages have a similar structural basis. Different languages still exist because there are certain parameters in the Universal Grammar which can still be set.

In the generative grammar view, the brain is modularly organized. The syntax generator functions completely separately from the lexicon and from the conceptual system. Syntax, word meanings and concepts are therefore independent.

The main argument for generative grammar is the Poverty of Stimulus argument. It claims that children are not exposed to enough linguistic data to be able to effectively learn a grammar. Children only get positive evidence for complex procedures such as forming questions from statements. Yet, all children manage to learn language, including knowing what ungrammatical questions are, in a very short time span. A universal grammar must be necessary to explain why children manage to learn a language, even without negative feedback.

Another argument for generative grammar is that all languages seem to be very similar. Grammatically poor languages such as pidgins eventually acquire rules and native speakers. By changing inflections, a statement can be transformed into a question. Each language uses multiple negation for denial.

One could wonder why children are not immediately fully proficient in their utterances if their grammar is already in place at birth. The explanation given by generative grammarians is that children do not yet have the attention span or processing power to form sentences. According to Chomsky, the children do have the *competence* to speak a language, merely not the *performance*.

However, there are still some issues to be had with generative grammar. Neurobiologically, there is no evidence for the existence of an autonomous syntax producing module. In fact, there is no neurobiological evidence for the possibility of any autonomous modules at all, the brain's structure is intricately interconnected.

Furthermore, the Poverty of Stimulus argument has never been empirically proven. The empirical studies that are being done indicate that if an innate grammar would be in place, the productivity of the child is very low. Productivity is centered around specific items and verb constructions, and does not immediately generalize to all members of the same category. Syntactic rules are not implemented throughout a child's language, but are limited to certain groups of verbs, so called *verb islands*. Slowly the boundaries between the verb islands disappear, as the syntactic rules become more pervasive. If there would be a universal grammar, being able to learn one kind of verb phrase should result in the ability to learn all verb phrases, which appears not to be the case [9].

Finally, generative grammar does not include any links to semantics. The meanings found in the lexicon are carried to the syntax by means of *lexical insertion*, but there are no rules for the semantic interpretation of a sentence. A sentence must therefore always constitute the sum of its parts. In a formal language, this would be perfectly fine, but in natural language, this presents a problem with the interpretation of sentences that are perhaps metaphorical in nature, such as for example *hitting the hay* or commonalities of the vernacular such as *coming round*.

2.2. Construction grammar

Construction grammar emerged as a reaction to generative grammar and its implications for cognition. Not being able to explain many linguistic expressions by means of their syntax and lexical meanings was deemed unacceptable. The untouchable 'well formed sentences' and grammaticality were not considered essential for the comprehension of language. Construction grammarians therefore decided not to see grammar as an innate set of rules, but as a corpus of constructions, from which analogies can be made.

They also do not see syntax and semantics as autonomous entities, but as inherently intertwined components of linguistic phenomena [10].

The differences that construction grammarians put forward negate many of the objections made to generative grammar. Instead of the possession of an ideal grammar, language becomes the ability to produce and comprehend language. Grammaticality is no longer defined by a universal set of rules, but by the customs of those who use the language. The generation of incomprehensible deeply nested grammatical structures is no longer possible.

A less reactive origin of construction grammar lays in cognitive linguistics. Cognitive linguists believe that language creation, learning and usage is not far from other forms of creation, learning and usage in cognition in general. They aim to unify both neurobiological findings and cognitive science- the densely interconnected structure of the brain, the initially nonsensical seeming babblings of children and other phenomena should all find intuitively adequate explanations [2].

If grammar is contained within a corpus of constructions, what do these constructions look like? The constructions consist of two poles, the conceptual and the lexical. The conceptual pole consists of an image schema. The lexical pole consists of lexical or syntactic patterns as observed in the language. Both categories are prototypical– both can be learned from exemplars. Any new linguistic input therefore can alter the corpus.

A result of construction grammar is its implications for learning theory. Since syntax and semantics are inherently connected, both must be learned together. Language learning therefore consists of collecting constructions of varying (and increasing) degrees of complexity. We can easily identify this approach with the observably slow expansion of child utterances.

A child's language is strongly influenced by its surroundings, which offer linguistic input to keep in its corpus. Similarly, language itself is determined by those who speak it: grammaticality is determined by the behavior of a group of speakers (by their corpora) and not by some 'ideal grammar'. Construction grammarians see language as a dynamic system which is actively maintained by those who speak it.

3. Hypothesis

The arguments between the two opposing views on language remain mainly theoretical. To be able to determine whether either view is a possible answer to how language is stored and acquired, we would need actual evidence. If one could make a computational model which could acquire language in the same way the generative or construction grammarians postulate, then that would make that view a plausible one.

In this paper we present a prototype of an implementation for a computational model of language acquisition according to the views of construction grammar. If the computational model is successful, it would be able to learn syntactical constructions and semantics without using an innate grammar. Most importantly, it would be able to acquire the first corpus that a child would later use to find meanings for new linguistic input.

Can we make a computational model for the acquisition of a child's first corpus? We have attempted to make a model of performance based language acquisition through to two word constructions. We would like the model to be able to generate new two word constructions based on arbitrary situational input frames.

Section 4 gives an outline of how we perceive the steps of language acquisition using construction grammar. Section 5 will subsequently deal with how we have implemented this model. Section 6 shows some of the results obtained through this preliminary implementation, and finally Section 7 will explain how we think the implementation can be expanded.

4. Model Outline

Observing language acquisition in children leads to 4 rough stages: the babbling stage, the one word stage, the two word stage and the multi-word stage (see figure 1). Within these 4 stages, there are 6 main research topics: segmentation (knowing where words begin and end), how children learn the meanings of single words, semantic bootstrapping (finding the meaning of a word based on its syntactic category), child language (the strange grammatical makeup of the language that precedes adult language), distributional categorization (finding syntactical categories to assign words to) and syntactic bootstrapping (the

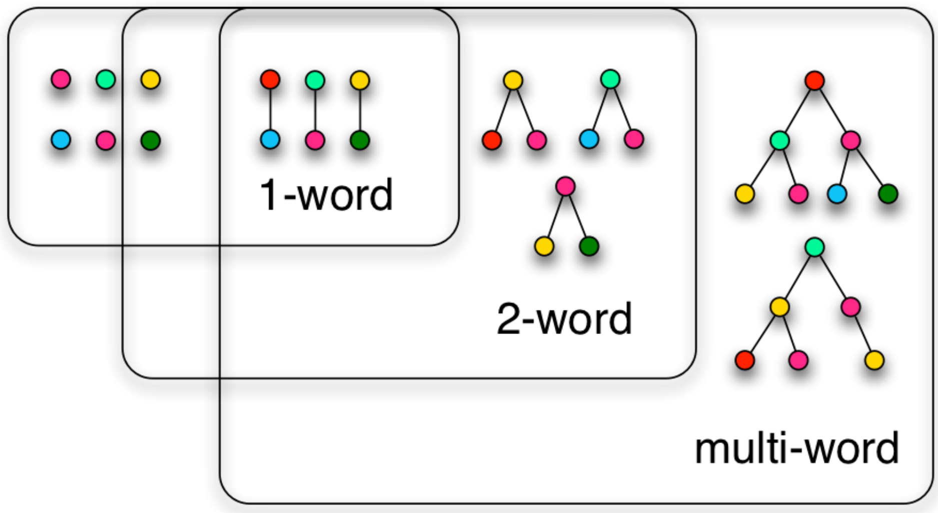

Figure 1: From the prelinguistic stage to the multi-word stage.

distribution of the word cues its meaning) [6].

Our computational model will be loosely based on the model of grammar learning as described by Chang and Maia in [3] and as interpreted by De Kreek [6]. In their model they make several assumptions, which we will list below, followed by an outline of the model.

Chang and Maia assume that before the process of language acquisition starts, children do have the conceptual ability to observe and discern situations and processes. This is fairly intuitive, if one would consider the prelinguistic stage as analogous to the cognitive state of animals. There is a general understanding of a situation, but as of yet no symbolic representation. Chang and Maia also assume that children have the ability to segment acoustic signals into words and memorize them together with the situational frames they occur in. Memories are then built up out of sound sequences which are connected to situational frames in a convergent process.

This segmentation process is similar to learning words, and is not dealt with either in Chang and Maia's technical model. They do however, make two important assumptions about word learning. First, they assume an early one word stage where the uttered words are connected to very specific actions, contexts or events. Second, they assume that even if a child is using the same word for two separate events, that does not mean the child has learned the full meaning of the word, but perhaps that there are two separate connections in the child's corpus which both happen to contain the same sound sequence.

In [8], Maia and Chang offer an interpretation of semantic bootstrapping in two parts. First, the prelinguistic conceptual categories become the preliminary syntactic word categories. Second, the prelinguistic representation of situational frames become the first linguistic constructions.

Chang and Maia also allow the generation of lexical constructions found by mapping linguistic forms on to prelinguistic action frames. This leads to simple transitive verb constructions such as *throw ball*. Once a critical amount of linguistic forms have been mapped to prelinguistic action frames, Chang and Maia think that the child will become more inclined to learn more words, because it has become an expectation. After this critical point, the child will start being able to generalize over transitive verb constructions, creating the so called *verb islands* in basic child grammar.

Distributional cues possibly could continue the generalization process. Non-physical action verbs could become part of the same group as physical action verbs, allowing even more general syntactic categories.

Syntactic bootstrapping is not considered by Chang and Maia. However, a similar process seems to be assumed, because the amount of prelinguistic situational frames that can be learned is fairly limited. Because of this, some form of syntactic bootstrapping would have to take place to be able to learn about new situations.

The main focus of Chang and Maia's model is the finding of basic child grammar. The assumptions they make about the prelinguistic stage strongly influence the outcome of the model. Using their model means that we would also have to agree with their assumtions and realize that the way we simulate the prelinguistic stage can be crucial to the rest of the language acquisition model.

4.1. Form of prelinguistic structures

The prelinguistic structures are given form through situational frames. Each frame can comprise roles, entities, utterances, operators, and natural categories. Before an action frame such as 'hold' is known to have roles for the 'holder' and the 'held', it can be considered as a stand alone concept. Later the action can be mapped to an entity through a role.

```
ACTION: HOLD
    ROLE: HELD[baby]
    ROLE: HOLDER[mother]
```

After the action frame has been generated for specific entities such as 'baby' and 'mother', it can be generalized for a whole category of entities. According to Chang and Maia, 'biologically natural categories' will become the initial syntactical categories. Therefore, the child will begin by abstracting to categories such as *physical action, human, physical object, etc.* at first.

```
ACTION: HOLD
    ROLE: HELD[human]
    ROLE: HOLDER[human]
```

The generalizations can continue to levels within the frame, resulting in completely abstract relational frames.

```
ACTION: DIRECTED
    ROLE: ACTOR[human]
    ROLE: OBJECT[human]
```

However, since there are only a limited amount of biologically natural categories, there are also only a limitied amount of abstract relational frames according to Chang and Maia. This is why child grammar differs from adult grammar.

4.2. The One Word Phase

Once the prelinguistic phase is in place, the one word phase is only a statistical analysis of the occurrence of words within situational frames. A situational frame is split into all possible subframes (including the situational frame itself) and all words in the utterances are connected to the available subframes. Each time a word is encountered together with a particular subframe, that word-meaning combination is reinforced.

Also during the one word phase, a child will be seeing situational frames without paired utterances. For each subframe of the observed frame, all associated words will be found and their corresponding appearance value will be increased by the value of that word in that frame.

The one word phase thus results in a lexicon with associated situational subframes ("meanings"), words, and their appearance scores.

4.3. Grammatical Constructions in the Two Word Phase

In [3] Chang and Maia propose a model for the example based learning of the first multi-word grammatical constructions. These constructions are taken to be mappings between form (in this case word order) and meaning. The examples consist of an utterance paired with a situation represented as a set of situational frames. Assumed here is that a correlation is expected between what is heard and what is perceived.

This model is based on three interlocked processes: analysis, hypothesis and reorganization. The *analysis* process determines which constructions are relevant given an utterance and a situation and selects the best fitting subset out of these. The *hypothesis* process then tries to account for any data that has not yet been explained by the constructions found during analysis by forming new constructions. Working beside these processes, *reorganization* tries to assemble new constructions in a more bottom-up way. Namely by searching for similar or co-occurring constructions of which generalizations can be made. In the following sections these processes will be explained in more detail and our views on how they should be implemented will be explained.

4.3.1. Analysis

This process produces the best fitting analysis of a situation, consisting of constructions whose meaning and form constraints are met by the situational frame and the perceived utterance respectively. For example, given the utterance "throw the block to mama" a set of known words, e.g. {*throw*, *block*, *mama*}, is formed and their lexical and syntactic constructions are cued. These constructions are then matched against the utterance and situational frame to produce a set of analyses. These analyses have an associated cost, corresponding to how well they account for the data in the situation. The best-fitting analysis, that is the analysis with the lowest cost, is subsequently rewarded and the cued meanings that where not used to produce this analysis are penalized.

In our application of this model analysis cues semantic constructions corresponding to the known words in the perceived utterance. These are the structures constructed during the one word stage. Next these structures are fitted to linguistic structures with a variable place in both the meaning frame and the word order, completing these construction. These are fitted to the situational frame and utterance. This is more focused on the specific structures of two word utterances in that the structures only allow room for the lexical binding of two words, no more.

4.3.2. Hypothesis

New linguistic structures are made from data that was unaccounted for in the analysis of a situation. Syntactic and lexical structures are extracted from this data. From these structures a potential construction is somehow made, explaining the previously unexplained data. Pairs of form-meaning bindings that are constructionally related are then used to reanalyze the situation. If the cost of this analysis is lower then the initial analysis then the new construction is admitted into the the corpus.

The process that hypothesizes new constructions in our adaption of the model uses a form-meaning construction that is specifically adapted to be used for the two word phase, but which could be recursively used to explain multi-word data. This construction has a variable place in both the syntactic frame (word order) and the semantical frame. Only if the situation cannot be analyzed using multi word structures will new constructions be hypothesized. This is done by searching for pairs of lexical frames corresponding to the words in the utterance of which one is a subframe of the other. This part of the other frame will be made variable, making a new two word construction. All the made constructions will be admitted into the corpus, but it is expected that bad constructions will not be reinforced later on.

4.3.3. Reorganization

In Chang and Maia's model a reorganization process is proposed that constructs new form-meaning mappings using a bottom up approach. This constructs these mappings by searching the corpus for sets of constructions that share similarities or that have a high frequency of co-occcurance. This process merges constructions involving correlated relational mappings over pairs of constituents that can be generalized using a conceptual ontology. Because of the explorative nature of this project this process has not been included in our model. The point of this project is to show that new two word constructions can be bootstrapped from single word lexical frames. This is already done by analyzing situations followed by the hypothesizing of new constructions out of unexplained data. However, we do feel that this decision could have effects on the results, but that it is not a necessary condition for two word structures to be acquired.

4.4. The Multi-word Stage, Adult Grammar and Data-Oriented Processing

After the two word stage is in place, the steps to continue towards a multi-word stage become much easier. A mechanism which takes all available constructions (lexical or larger) into consideration for making a generalization will start producing larger constructions, which will then be added to the corpus. Once larger constructions are available for new generalizations, the corpus will rapidly expand with many multi-word constructions.

A mechanism as described above already exists with merit for adult language processing. It is known as Data-Oriented Processing and was introduced by Remko Scha in 1990 [1]. The analysis algorithm as described above is really also a linguistic component of the Data-Oriented Processing framework.

Once the two word stage has been implemented, a logical step would be to combine it with Data-oriented Processing. The theoretical implications of such a combination have been described by de Kreek in [6], but have not yet been implemented for this paper.

---

5. Implementation

The three phases of our model –corpus annotation, one word phase and two word phase– are implemented in seperate but dependent modules. First we needed a representation of the input corpus that mimics the prelinguistic stages. We chose to use an `xml` format, in the hope of making this data generally accessible through readily available libraries. Second, we implemented the one word algorithm which maps words to linguistic meanings in Python, together with methods to access and compare the input data. Finally, for the two word phase, our formal specifications for learning linguistic abstractions were implemented, building on top of the output and methods of the one word phase. The annotation of the input corpus and the one word stage have been inspired by work done by Van Santen in his Bachelor's thesis [11]. His thesis and our two word specification were based on the theoretical approach of De Kreek [6].

5.1. Corpus Annotation

Within our project, we have three different datasets. These are the CHILDES corpus [4], our annotated corpus and the linguistic corpus. The CHILDES corpus consists of utterances taken from conversations between children and their parents. The annotated corpus was constructed from the CHILDES corpus and contains these utterances coupled with situational descriptions. These were the data used to train the one and two word stage. The linguistic corpus consists of the output of the one and two word stage. This output comprises word-meaning associations and derived linguistic abstractions and will be discussed in the one word and two word sections respectively.

5.1.1. CHILDES

The CHILDES database, or Child Language Data Exchange System, is a collection of encoded interactions of children. The data of importance to us are the situational descriptions combined with the adult utterances accompanying them (the adult is usually the mother). Example utterances are:

```
*MOT: more juice ?
*MOT: would you like more grape juice ?
*MOT: where's your cup ?
```

We made a selection of this comprehensive database based on the age of the child in question, roughly between the age of one and two, since that is the age when children are usually in the two word stage. This selection from the CHILDES corpus had to be manually annotated with semantic information.

5.1.2. Annotated Corpus

The annotated corpus is essentially a collection of independent situations. Each of these situations consists of a situational description and one or more adult utterances. For our model we assume the concepts in these descriptions to be available before the words or linguistic abstractions are learned, in line which Chang and Maia [3]. The situational descriptions in the annotated corpus have been constructed using descriptions in the CHILDES corpus, but they are mostly our own interpretations. The utterances on the other hand are taken literally from the CHILDES corpus.

Our corpus data is in `xml` format, but below we will give an example situation in a more readable form. The description element (1) is only for presentation purposes. The frame element (2) signifies the basic mental construct the child is supposed to have. Each frame has an id (3) and often an abstraction (4). It is important to note that these are used only to identify mental constructs. They are not textually compared to the utterances. A frame can also have subframes which fulfill specific linguistic roles, like the actor (5) and object of a situation. Finally, one or more adult utterances (6) are added. The exclamation mark in the utterance can be used to emphasize a word. Not shown here are properties. Properties can be added to frames and function roughly the same as real subframes.

```
(1) DESC: talk about juice
(2) FRAME: action
(3)         ID: want
(4)         ABSTR: desire
            FRAME: object
                    ID: grapejuice
                    ABSTR: object:food
(5)         FRAME: who
                    ID: child
                    ABSTR: object:human
(6) "more !juice"
```

The annotated corpus is the actual data we used as input to map words to "meanings" in the one word stage. These meanings are considered to be the semantic denotation of the words with which they are associated and consist of a frame like the one shown above, or a sub-/superframe of such a frame. [1]

The output of the one word stage is a data structure of associations between words and frames, along with scores. When the code is run a word can be looked up to see which meaning frames it is associated with, represented as the top five best scoring frames. This data structure will be used by the next stage.

5.2. One Word

To the one word model, the corpus is a collection of unrelated situations– each situation is processed seperately. The goal of the one word stage is to try to associate words with semantic concepts, by making an index of word-meaning associations, which in turn have a certain score. The two main procedures to fulfill this goal will be outlined next. Firstly, the manner of scoring is to be defined, as if it were an association between a word-meaning pair. In its simplest form, this consists of counting the number of times that a word is found together with a specific meaning, across all situations.

Intuitively, it seems that this straightforward cumulative score should be corrected for how frequent the word and/or meaning is, so that words that occur often will not bias the results[2]. However, after experimenting with dividing this score by either the total frequency of a word or of a meaning, it seems that both fail to consistently improve the results. More specifically, it resulted in words being associated with bigger, but not necessarily better, frames (eg. whole actions instead of object frames – often not appropriate when the word is an object), or it broke the two word stage in that it could not construct any linguistic abstractions anymore. This would be an important area for future research.

The second important function in the one word code consists of "deriving meanings" from an input situation. Which meanings are derived has mostly been a matter of choice. Based on intuition we decided to derive the following from the situational frame: the original frame; all the complete subframes; abstracted versions of all the frames found; all single properties. At this point, our implementation differs from the van Santen's implementation, which eg. might generate 90 vs. 10 subframes in our code, from a given situation. This choice is based on the idea that generating lots of similar subframes will not improve the results, but only increase computational complexity.

5.3. Two Word

In the two word stage linguistic structures consisting of a word with a variable slot for another word will be learned. These structures are called *linguistic abstractions*, because they are abstracted from previous language experience. These abstractions are made by analyzing the same annotated corpus that was used for learning in the one word stage. Each of the situations in this input is processed in turn. Once an abstraction is learned it is added to the linguistic corpus that is used to analyze encountered situations. If one of these abstractions later on turns out to be useful in analyzing a situation it is reinforced, making it more likely to be used again in analysis.

The two word stage uses the linguistic corpus that is constructed during the one word stage to achieve this. The linguistic corpus consists of a collection of words coupled to their semantic denotations, or meanings, which are represented by situational frames such as previously described. Multiple meanings can be coupled to a single word and they are ordered by their score, which is a measure of how strongly a meaning is mapped to a word.

When a situation is being processed it is first analyzed, meaning that already learned linguistic structures are sought that are suitable for this situation. This is done by first searching for possible meanings for each word in the utterance that accompanies the situational description. These meanings have to be among those that are bound most strongly to the word and they have to be compatible with the situation. This means that the meaning frame also has to be a subframe of the situational description. We chose to select the five highest scoring meanings that where found in order to assure that the found meanings where reasonably strongly bound to the words, while still retrieving enough meanings to ensure that a linguistic abstraction could be found. This list of meanings is then used to find abstractions for which one of these meanings fit into its variable place. This means that the meaning should have a value for ev-

[1] At the start of the project we received an example of a data corpus, on which we based our specific implementation.
[2] Cf. tf-idf weight (term frequency-inverse document frequency)

ery part of the abstraction that is variable and that the meaning does not contain any parts that are not found in the abstraction. If a set of these abstracted structures are found, they are reinforced. This is implemented as a simple counter, corresponding to the number of times that the structure has been found to fit while analyzing a situation. The linguistic abstractions reinforced most strongly are the ones that are most likely correct.

If none of these structures are found then new abstractions are made from the situation that is being processed. This is done in a succession of steps. First sets of two words whose meanings are connected are sought. With "connected" is meant that one of the meanings is a subframe of the other. Secondly new abstractions are made out of the higher level[3] meanings. This abstraction is done as following: if the lower level meaning consists of a property the value for the property will be made variable, if it is an object frame all properties and the id are made variable, but not the abstraction. This last abstraction corresponds to a role in an action frame that can be fulfilled by any member of a more abstract group.

Thirdly the LA is completed by adding the word order in which the two words appear in the utterance. The word that is coupled to the meaning from which the abstraction was made is preserved, while the word coupled with the subframe is made variable. A completed abstraction will then look something like this:

```
LINGUISTIC ABSTRACTION:
    FRAME: action
            ID: want
            ABSTR: desire
            FRAME: VAR
                    ID: VAR
                    ABSTR: object:food
                    PROP: VAR
            FRAME: who
                    ID: child
                    ABSTR: object:human

    WORDORDER: want BEFORE VAR
```

Using a corpus of these linguistic structures, new two word utterances can be produced by fitting them onto a presented situation. If they are compatible then the two word utterance will consist of the word coupled to the linguistic abstraction and the word coupled to the part of the situation frame that fits in the variable part of the abstraction frame. This last word is found in the part of the corpus produced during the one word stage. These two words are then presented in the order specified in the abstraction frame.

6. Results

Experiments with our implementation yielded some hopeful results, both in the one word stage and in the two word stage. At the very least, new linguistic constructions and two word utterances were formed by combining the annotated corpus, the one word associations and the two word abstractions. Analyzing the output however is not completely trivial. How much of the constructions and accompanying two word pairs are really meaningful is subject to discussion.

6.1. One Word

In general the one word algorithm found satisfying results, at least conceptually. Words describing objects were associated more often than not with frames representing those objects. One prevailing error was the strong association of single properties with words, for example "block" being something square, rather than a complete block-object frame.

```
block

match 1 score: 31
MEANING:
        PROP: shape = square
------------------------------------
match 2 score: 29
MEANING:
        FRAME: object
                ID: object:toy
                PROP: shape = square
------------------------------------
match 3 score: 25
MEANING:
        FRAME: object
                ID: block
                ABSTR: object:toy
                PROP: shape = square
------------------------------------
```

Verbs, usually representing action frames, presented less desirable results. Often they were associated

[3] A frame is considered to be of a higher level then another frame if the other frame is a subframe of the first.

more strongly with objects than with complete actions. This does not inhibit the two word stage from working, as the five frames showing the strongest association with a word are used by the two word stage. These five frames always contained at least some useful results.

6.2. Two Word

We used two test-cases for the two word stage. One with a small, specifically crafted corpus and the other with real world data extracted from the CHILDES corpus.

The handcrafted corpus made it possible to predict the outcome and compare it to the actual output. In this case the two word stage gave the desired results. This proves that at least our algorithm functions according to specifications.

The results of the real world corpus were much harder to interpret. It contains hopeful output, and does indeed create novel utterances. Often however, it leaves one to wonder whether an utterance is really meaningful, or just some artefact of the input data. For example, the error from the one word stage with single properties gave rise to some strange linuistic abstractions, eg. "throw" might have an abstraction specifying that it is an action of throwing a block with a variable shape. The error with action frames being associated with objects resulted in non-sequitors like "block block." In the example output below, we can see that word combinations can appear meaningful to us, although sometimes grammatically incorrect.

```
Situation: 24
1 : ball give   Score = 209
2 : give ball   Score = 129
--------------------------------------
Situation: 108
1 : shut door   Score = 13
2 : shut door   Score = 7
--------------------------------------
Situation: 157
1 : hold you    Score = 29
2 : hold still  Score = 29
```

In situation 24, both "give ball" and its reverse are generated. The association is all the more clear, given that neither utterance existed in the annotated corpus. Why a given word order is more likely than another is probably an artefact of this input corpus. In situation 108, it is shown that the same word combination can be generated twice by using different linguistic abstractions. Situation 157 examplifies an oddity, where "you" is treated as an object. Looking at the corpus, this is actually quite intuive since "you" is most often said in combination with the presence of the child. The one word output justifies this intuition. "Hold still" on the other hand, was the actual utterance spoken in that situation, which shows that the model is also able to reconstruct meaningful utterances.

The most important question, whether our implementation could provide a corpus of linguistic contructions able to generate novel utterances can be answered in the affirmative. Roughly 58% of the utterances generated did not yet exist as such in the annotated corpus.

7. Conclusion

In this project we examined the merit of a *construction grammer* approach to language acquisition. By implementing a computational model based on the theoretic assumptions of this approach, we wanted to demonstrate that *construction grammer* is a plausible alternative to the generative grammar approach. In the hypothesis we stated that the model should be able to derive a corpus of linguistic constructions that facilitates novel utterances. We found proof that this is indeed the case. The model c.q. the child showed promising output, where new word pairs were created based solely on linguistic abstractions. This shows that in principle, performance based language acquisition is possible.

This is however just a prototype implementation. There are questions still unanswered, and some assumptions are not yet validated. In future work, a few specific points should be addressed. First the corpus annotation is based merely on an intuition of how a child incorporates mental constructs. There might be better ways to represent this. Second the scoring algorithms are very simplistic. A more in depth statistical approach could improve the results, for example by acknowledging that situations are not unrelated but occur in time. Last, our implementation is not really exemplar based. Specifics about the complete utterences and actual situations are neglected altogether, focusing only on the derived linguistic construction. If these shortcomings can be resolved, we believe it is possible to create a convincing model for language acquisition.

8. Acknowledgments

This paper is one of the products of an AI bachelor second-year project at the University of Amsterdam. This project was supervised by Remko Scha, and we would like to thank him for all his help and support. We would also like to thank Mike de Kreek, Mart van Santen and Gideon Borensztajn for all of their previous work, and Marco Wessel for hosting our svn server.

References

[1] R. Bod and R. Scha. *Data-Oriented Language Processing: An Overview.* ILLC, Universiteit van Amsterdam, 1996.

[2] G. Borensztajn. Automatic discovery of construction in children's speech. Master's thesis, Universiteit van Amsterdam, 2006.

[3] N. C. Chang and T. V. Maia. Grounded learning of grammatical constructions. In *2001 AAAI Spring Symposium on Learning Grounded Representations*, 2001.

[4] CHILDES. http://childes.psy.cmu.edu.

[5] N. Chomsky. *Aspects of the Theory of Syntax.* The MIT Press, Cambridge MA, 1965.

[6] M. de Kreek. Language acquisition and virtual grammars: On the continuity of cognition. Master's thesis, Universiteit van Amsterdam, 2003.

[7] E. R. Kandel, J. H. Schwarz, and T. M. Jessel. *Principles of Neural Science.* McGraw-Hill, 4. edition, 2000.

[8] T. V. Maia and N. C. Chang. Grounding the acquisition of grammar in sensorimotor representations. In *2001 AAAI Spring Symposium on Learning Grounded Representations*, 2001.

[9] M. Tomasello. *First Verbs: A Case Study of Early Grammatical Development.* Cambridge University Press, 1992.

[10] M. Tomasello. *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Harvard University Press, 2003.

[11] M. van Santen. Bootstrapping Language. Bachelor's thesis, Universiteit van Amsterdam, 2005.