Simulating Language Games of the Two Word Stage using Unification and Substitutions on a Corpus of Exemplars with a focus on Semantics

. . . being an endeavor in cognitive simulation to parsimoniously re-enact verbal interactions of a toddler through translation and reckoning with pragmatic and semantic annotations of its linguistic history.

*Andreas van Cranenburgh**

June 8, 2009

Bachelor thesis Artificial Intelligence: second draft
Supervisors: Remko Scha & Jos de Bruin,
University of Amsterdam

Try to attach a meaning
To words that you've heard

Stumbling through the dark
Seems I'm stumbling through the dark
Eveybody's stumbling through the dark

The men who proceeded us here
Left only questions and fears

[...]

— from the album *Rainy day music*, The Jayhawks (2003)

*acranenb@science.uva.nl, 0440949

Contents

1 Introduction

General linguistics has been dominated by Chomskian generative linguistics for several decades. The focus is on rules and their creativity, viz. systematicity and productivity. The central dogma is that an in-born, Universal Grammar is necessary to adequately explain these phenomena. It holds on to the continuity assumption, which states that language as used and understood by children is qualitatively equal to that of adults (for criticism, cf., Tomasello (2005)).

However, from a developmental psychology angle, several empirical findings (Tomasello, 2000; 2005) shed doubt on whether this approach is applicable to language acquisition by children. It rather appears that language learning is bootstrapped in a haphazard fashion, learning constructions here and there, which can only later be synthesized to form a coherent grammar.

Rather than trying to resolve this age-old debate between rationalism and empiricism along theoretical lines, it might be fruitful to try to model the behavior of early language users, and demonstrate in this way that a universal grammar is in fact not necessary to explain the phenomena observed. This strategy echoes a suggestion made by Turing (1950):

> "Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? [...] Presumably the child-brain is something like a notebook as one buys it from the stationers. Rather little mechanism, and lots of blank sheets."

2 Theory

2.1 Literature review

One of the foremost proponents of the developmental take on language acquisition is Tomasello (2005). He argues that linguistic abilities are acquired gradually, in an incremental fashion. Linguistic forms are memorized in conjunction with their communicative functions or meanings. These constructions are then generalized so that language use becomes ever more expressive and productive. Aspects which distinguish this approach from that of generative linguistics is the rejection of the autonomy of syntax and the consequential focus on semantic and pragmatic influences on learning. Aside from that the idiomatic and figurative dimension of language presents problems for purely formal accounts of semantics and syntax.

The formal nature of traditional theories goes back to American (Bloomfeldian) structuralism and the supposed arbitrariness of the sign. A Counterargument to the arbitrariness of the sign is that derived (eg., figurative) meanings are relatively systematically related to their canonical meanings. For example, the verb 'to come' has the canonical interpretation of spatial movement, but it can also be applied to an event which is temporally approaching: "Christmas is coming" (Lakoff and Johnson, 1999). Notice how 'approaching' is also a spatial verb, and can be analogously applied with a temporal interpretation.

The work of eg. van Kampen (2003) on children's' use of languages in the two word stage indicates that their (proto-)grammar employs pragmatic operators and content signs, instead of distinguishing all the syntactic categories present in adult language. Verbs are not yet inflected, and determiners are absent.

Chang and Gurevich (2004) demonstrate a computational model of Embodied Construction Grammar that combines constructions to interpret new constructions. Their semantic representation could serve as an inspiration. Also, the use of Minimum Description Length learning provides a good way to prune the database of learned constructions.

Steels (2004) describes his experiments with situated agents (robots fitted with cameras) that employ language games as a learning strategy. An example of a language game is the description game: one agent describes an event that has just happened, and the other responds by agreeing if the description matches its own experience. These experiments simulate language

```
1. "ball gone"  la score = 1
LINGUISTIC ABSTRACTION:
        WORDORDER: VAR:gone
        FRAME: action
                ID: action:move
                FRAME: object
                        ID: VAR
                        ABSTR: object:toy
```

Figure 1: Example situation in a previous model

genesis and grammaticalization *ab inito*.

van Kampen and Scha (2007) discuss the modeling of early syntax acquisition using the Data Oriented Parsing framework (Bod and Scha, 1996). This means that all input is stored in memory, and made available for recombination in the recognition of novel utterances.

2.2 Motivation

A previous project (van Cranenburgh et al., 2007) attempted to model the acquisition of constructions in the two word stage of early child language. The model used a corpus of utterances spoken to children, annotated with semantic representations of the context. The aim was for this model to be able to generalize over the sentences to discover the correct associations between words and their semantic representations, and to be able to combine sentence fragments into novel utterances. This model did not consider syntax and semantics separately, in the style of construction grammar (as employed in eg., (Tomasello, 2000; 2005)). Although indeed correct associations were found, and novel utterances could be recognized, most of the former were incorrect, and most of the latter non-sensical (although in part this was due to the first issue worsening the second). Figure 1 illustrates an example of an utterance as it was interpreted (in this case correctly) by this model.

In this sentence the construction "X gone" was applied to "ball", because it matched the condition of being a toy. The construction was apparently previously encountered when a toy was being moved. In this case the result was satisfactory, but unfortunately most other abstractions were spurious.

The problem was that sentences were being learned as isolated fragments, without any notion of discourse or pragmatics. Also, the semantic representation did not fit well with all the words to be learned: it was only good at representing actions and objects; prepositions and demonstratives and other abstract words were not being learned. Instead of merely focusing on semantically describing a situation, the learner should consider the total communicative function of an utterance. The learning was implemented as making associations between words and each part of the semantic representation, and counting how often these associations occurred. This meant that a lot of incorrect associations were made. Unfortunately the model did not make use of pruning, as there was no way to know which associations had been incorrect.

Last year another project (Odolphi, 2008) developed a formal grammar for the two word stage, based on empirical work on child language (eg., van Kampen (2003)). This grammar does not make use of adult-like syntactic categories such as verb and noun, but groups expressions as topics, comments and operators. Using this grammar it is possible to parse and produce child utterances, because it turns out that almost all of the two word utterances follow the pattern of this formal grammar.

These projects focussed on children's own utterances. However, it appears that children can already comprehend more complicated sentences than they produce themselves, as evidenced by such exchanges as:

```
*MOT:   wanna [: want to] put (th)em [= crayons]
        back in the box ?
%act:   <4-7> MOT taps the box with her finger
%gpx:   MOT looks at CHI
*CHI:   no .
%gpx:   <bef> CHI looks up at the box .
        CHI looks down at the chair
```

— Childes,[1] New England corpus[2], Liam, November 30th, 1984

2.3 Research question

Can an exemplar-based model of language acquisition account for the discrepancy between language comprehension and production of children in the two word stage? Can this model facilitate the simulation of simple language games of parent and child?

These questions will be addressed by attempting to implement a model of linguistic comprehension and production using an exmplar-based model of language.

2.4 Some philosophical considerations

2.4.1 The importance of philosophy for cognition

Since the cognitive revolution cognition has been conceived as symbol manipulation. This idea has kept researchers in both cognitive psychology and

[1] MacWhinney and Snow (1995)

[2] `http://childes.psy.cmu.edu/data/Eng-USA/NewEngland.zip`

artificial intelligence (AI) in business. The more or less official ideology of 'Methodological Solipsism' (Fodor, 1980) secures research grants by asserting that other fields such as neuroscience and biology can have no bearing on the subjects of the so-called 'special sciences.'[3]

This doctrine has come under fire from different directions. Within cognitive science itself there is talk of a second generation (Lakoff and Johnson, 1999) putting forth embodiment as vital; as well as a revival of connectionist systems with subsymbolic, distributed representations. But long before that there has been vocal criticism from philosophy. Dreyfus (1972) correctly predicted the failure of the bombastic ambitions of early AI, and basically claimed that this was due to the symbol manipulation metaphor being a pipe dream:

> "Philosophers have thought of man as a contemplative mind passively receiving data about the world and then ordering the elements." – Dreyfus (1972)

In short, Dreyfus warned that artificial intelligence is rather like alchemy: suffering from unwarranted optimism and badly in need of re-evaluating its dogmas. The last part is exactly what I am about to do.

2.4.2 Compositionality

Compositionality[4] is possible, but not necessary, in this model. In the most specific case, a whole sentence is fully described by a single exemplar; in the most general case, a sentence is interpreted word for word, one exemplar each. However, it would appear to be optimal to employ a sort of 'basic-level constructions,' (cf. basic-level categories; Rosch et al. (1976)) corresponding to stable collocations that describe a large number of sentences using a small number of multi-word fragments from exemplars. This should be optimal because it reduces the memory load, since not every sentence has to be stored, and because it allows the meaning of words to be dependent on sentence context.

2.4.3 Autonomy of syntax from semantics

The dominant trend in linguistics is syntacto-centrism, as observed by Jackendoff (1983). What makes syntax so interesting is rarely made explicit, but a desire for immediate and rigorous results probably favors the systematic nature of syntax, at the expense of the more elusive and sometimes vague

[3]Upon hearing the term 'special sciences,' I can't help but think of 'special' as in physically challenged and the image of 'physics envy' this evokes.

[4] "The meaning of a complex expression is a function of the meanings of its immediate syntactic parts and the way in which they are combined."
— Krifka (1999)

nature of semantics. Even accounts that explicitly focus on semantics and pragmatics are often syntactic in nature; a case in point is formal semantics. But whether these accounts actually describe semantics or merely mimic parts of it is a difficult question. Obviously planets do not need to be able to solve differential equations in order to orbit as we have come to expect. What we can be reasonably sure of is that semantics is residing (or perhaps presiding) in the human brain, but this is rather like predicting that it will rain somewhere, tomorrow. The useful question is whether there is some higher-order abstraction of semantics, and whether it can be mechanized or otherwise reproduced in certain systems. This entails that cognition is not just a projection or construction, but a valid abstraction over neural (and possibly other) details. This thesis makes the assumption that such an abstraction should exist, and that an approximation can be attempted and evaluated in a model of language use.

2.4.4 Mentalism

Most accounts of cognition and language in particular are mentalistic. That is, they posit a mental entity which manipulates explicit representations corresponding to external states of affairs. Representations, however, are problematic, because representations have to come from somewhere, either learned or innate, and should, serendipitously or otherwise, faithfully describe, ie., be isomorphic to, both distal external events and subjective experience. The most dramatic example is the 'Language of Thought' hypothesis (Fodor, 1975), which posits that cognition must operate on first-order logic predicates, which are taken to be universal and in-born. These predicates form the so-called semantic primitives, which can be composed to give rise to an apparent infinity of meanings – including such artefacts as door knobs and scissors which were certainly not part of our humble ancestors' inventories. It is safe to say such theories are far from parsimonious or even empirically responsible.

The other extreme is to reject mentalism and representations altogether, and stress the tight coupling of embodied agents with their surroundings; in a certain sense representations are made redundant by direct interactio with the world. In this conception there should be no need for the category of mind, and body and world are both necessary and sufficient ingredients for cognition. Rorty (1979) is a proponent of this view, and accuses the representationalist school of presuming language and mental events to be a "mirror of nature." Instead, he argues, there is no need for a mirror, and language is merely a part of nature. The later Wittgenstein (1953) rejected mentalism on the grounds that a private (mental) language would be impossible, because language is a social phenomenon, useful only by virtue of being shared and understood by a speech community. He argued that language games are the fundamental building blocks of language, in which language use is the sole criterion of meaning. In general anti-mentalistic accounts see language as a way of skillful coping with the world and one's con-specifics, as opposed to the possibly conscious manipulation of explicit symbols. Language is not a conduit which encodes propositional and illocutionary content, but a tool by which we negotiate our ways in the world.

Although I highly sympathize with these views it is highly difficult to apply such a philosophy to artificial intelligence, because it rejects abstract mental processes and representations *tout court*. This excludes the possibility of modeling aspects of language in isolation, since the situatedness of embodied agents is what cognition revolves about. Short of making a robot that catches up with millions of years of evolution, it would be impossible to responsibly model the cognition of language. To break this impasse I will make use of semantic and pragmatic representations, but without assuming them to be canonical and actually present in the minds of children. Instead they stand for or hint at experience and the spreading activation of neurons, and the social conventions immanent in language use.

3 Practice

3.1 Exemplars and Semantics

The model works with a set of exemplars. Exemplars contain an utterance and its meaning representation. The meaning representation consists of a speech act operator and a series of clauses, ordered by salience. Clauses consist of two predicates, where the first describes an action or category, which is predicated on the second. The first clause will often contain a topic and a comment, while the rest might contain context, presuppositions and associated facts salient in the relevant situation:

```
"utterance"
operator: pred1(pred2) pred3(pred4) ...
```

For example:

```
"what does a bunny do ?"
whquestion: do(X) animal(bunny)
```

This could be translated to logical form in the following way:

$$\exists P\, \forall x : [P(x) \rightarrow do(x) \wedge bunny(x) \rightarrow animal(x)]$$

where $P(x)$ is a variable predicate which should be instantiated upon answering the question. Another example:

```
"want some juice ?"
ynquestion: want(juice) food(juice)
```

This could be translated to logical form in the following way:

$$\forall x : [juice(x) \rightarrow want(x) \land juice(x) \rightarrow food(x)]$$

where $want(x)$ is a special predicate to represent intentional attitudes. But these translation do not preserve order, which is necessary because the expressed `want` is more salient than the implicit categorization `food`. This shortcoming of logical form is recognized in construction grammar, eg. 'the cat bites the dog' versus 'the dog is bitten by the cat' would have the exact same logical form but are nevertheless as distinct as can be.

Both predicates and their arguments can be variable by writing them as a single uppercase letter. Variable means that the information is missing from or asked in the utterance.

This representation makes no hard-and-fast distinction between what is explicitly verbalized in the utterance, and that which is understood through context, because this distinction would amount to a fully context-free, introspectable understanding of each and every word in the utterance. Instead of precisely describing the semantic structure of the utterance, this style of representation views the utterance as an ellipsis glossing over parts which can reasonably be expected to be filled in by hearers. Since this filling in of contextual details is not necessarily a linguistic phenomenon, it is assumed to have been completed successfully, and to be present in the initial corpus of exemplars.

3.2 Language Use and exemplars

[describe operations on exemplars here, without being specific about implementation]

Adequate participation in a discourse context requires interpreting an utterance, transforming this interpretation into an appropriate response, and verbalizing this response. Interpretation consists of finding a minimal covering set of exemplars which are compatible under unification or constrained argument substitution. Response generation is finding a best fit exemplar according to an operator to operator mapping (eg., question $\rightarrow$ answer). Verbalization is the mapping of instantiated clauses to lexical items inferred from multiple exemplar occurrences. Reinforcement (eg. when a parent reacts with "that's right") records an identifier linking the exemplars for the previous utterance and its response to strengthen their association.

3.3 The model

The first step in interpreting a novel utterance is finding the exemplar whose utterance is most similar to it. This is implemented by iterating over the ordered subsets of words occurring in a sentence, from long to short, and trying to find an exemplar containing these words. The meaning of the exemplar that is found is then used as a template to which other exemplars must conform if they are to be used in interpreting the rest of the utterance. An exemplar conforms to the current interpretation if it has a family resemblance with it, ie., one of its clauses has a predicate in common with the current interpretation. If the matching clause has a variable argument, it is instantiated. If it has a conflicting argument, it is substituted. In order to curtail spurious instantiations and substitutions, only clauses describing the words being covered are considered open to modification.

This requires some knowledge of the connections between clauses and words. This lexical knowledge is derived from the corpus of exemplars by juxtaposing all exemplars containing a specific word, and picking the most salient clause they have in common as the meaning for that word, or looking for links between clauses and words (explicitly, in the form of word indices, or implicitly, when words and predicates or arguments coincide). This process is repeated until no new definitions can be gleaned from the corpus of exemplars. Content words are especially likely to receive correct definitions from this process. This bias is acceptable because they are already acquired in the one word stage, as opposed to function words. Function words do not necessarily carry meaning in isolation, but rather co-ordinate and decorate sentence meaning, which is adequately contained in exemplars.

After finding the first exemplar further exemplars are sought in order to cover the remaining words in the utterance. The words are covered in a greedy fashion, the longest matching construction is used first. This process currently does not perform backtracking, but this could be added for cases where first matching a shorter construction enables a longer construction to be used later. See figure 2 and 3 for depictions of the steps involved with interpreting an utterance.

It is possible to use demonstratives in sentences, as long as the referent is supplied on input:

```
utterance: what is this [=ball] ?
meaning: whquestion: point(X) Y(X)
interpretation: whquestion: point(ball) toy(ball)
```

Interpreting this exemplar will cause the meaning of "ball" to be inserted, in this case by substituting the variable predicate Y with the meaning of "ball" (without a variable predicate it will be concatenated).

One of the advantages of the algorithm just described is its graceful degradation. Given sufficient redundancy, words can be misperceived or left out, and the remaining words might still enable correct interpretation. This feature enables natural interpretation of ellipses without specialized mechanisms:

```
Parent: kitty do ?
```

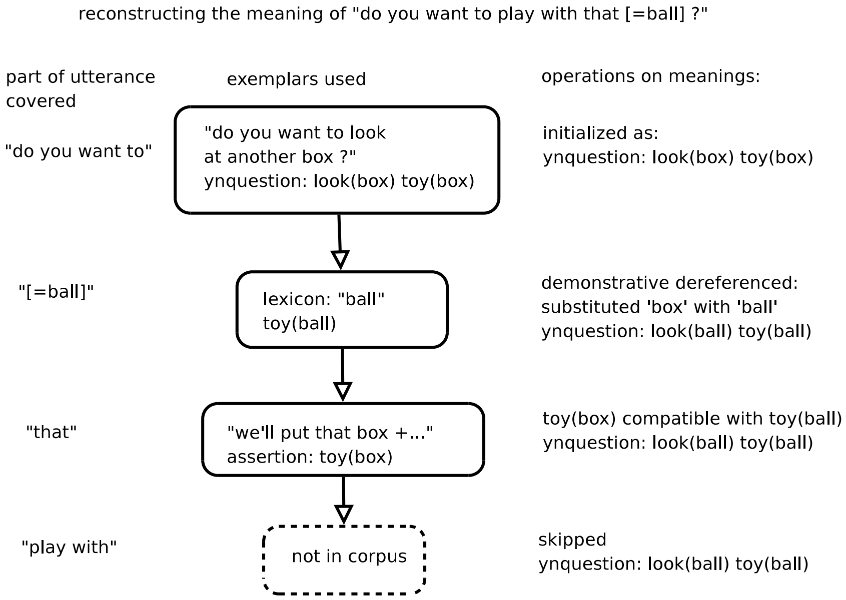

Figure 2: Interpretation process in a step-wise fashion

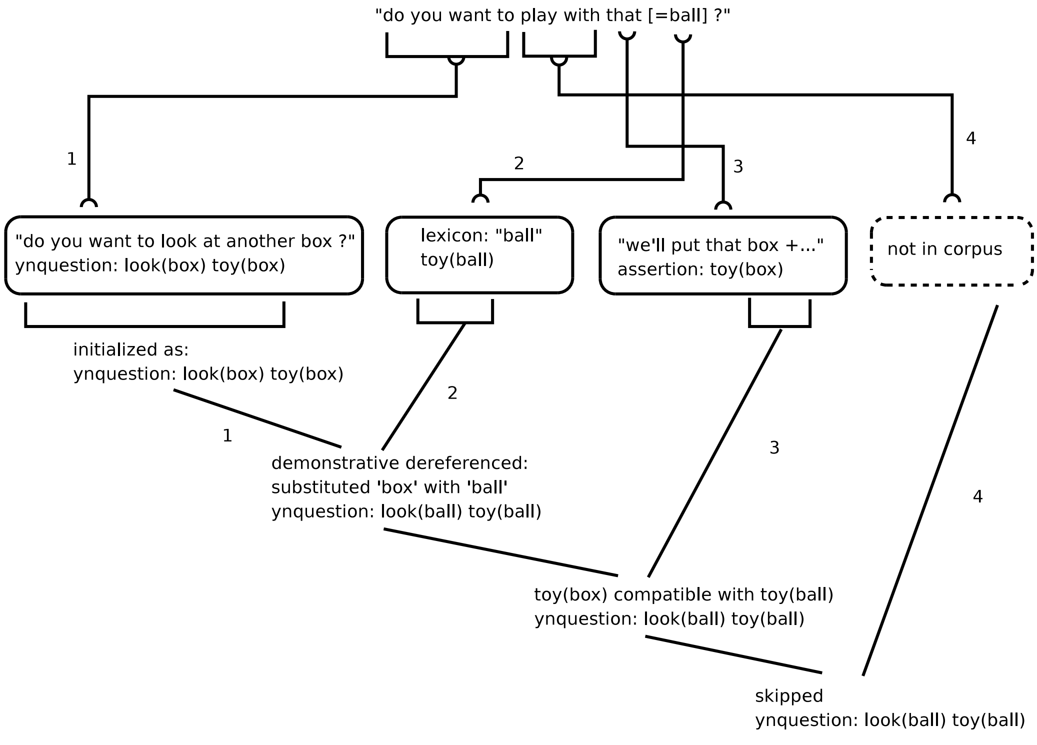

Figure 3: Interpretation depicted as resolution process

```
*MOT:   what shall we do ?
*CHI:   eat.
*MOT:   shall we eat cookies ?
*CHI:   ah
*MOT:   shall we ?
*CHI:   mmhmm.
*MOT:   where are the cookies ?
*CHI:   in bag.
*MOT:   cookie in the bag ?
*CHI:   baby eat.
*MOT:   ..
*CHI:   baby eat.
*MOT:   ..
*CHI:   cookies.
*MOT:   baby eat cookies ?
*CHI:   eat cookies.
```

(a) Childes fragment, as used in van Turnhout (2007)

```
*MOT:   what shall we do ?
*CHI:   eat
*MOT:   shall we eat cookies ?
*CHI:   0
*MOT:   shall we ?
*CHI:   0
*MOT:   where are the cookies ?
*CHI:   cookie bag
*MOT:   cookie in the bag ?
*CHI:   cookie bag
*MOT:   ..
*CHI:
*MOT:   ..
*CHI:
*MOT:   baby eat cookies ?
*CHI:   eat
```

(b) Model output

Figure 4: Comparison of Childes data and the responses generated by the model under discussion

```
interpretation: whquestion: do(X) animal(cat)
reaction: assertion: do(meow) animal(cat)
Child:  meow@o
```

3.4 Results

TBD

- show some example output (generalizations, ellipses)
- comparisons with childes data: see figure 3.4 and 3.4
- human judgments of output (double blind)

4 Discussion

TBD.

- Formulate the possible contribution of this work to the field of language acquisition.

```
*MOT:   that's the cow .
*MOT:   what's this ?
*CHI:   yyy .
*MOT:   is that a donkey ?
*CHI:   donkey .
*MOT:   right .
*MOT:   that's a donkey .
*CHI:   0 .
*MOT:   what's this ?
*CHI:   duck .
*MOT:   what does a duckie say ?
*CHI:   0 [<] .
*CHI:   quack@o .
```

(a) Childes fragment, New England corpus, Cristopher, July 19th, 1984

```
*MOT:   that's the cow .
*CHI:   cow
*MOT:   what's this ?
*CHI:
*MOT:   is that a donkey ?
*CHI:   donkey
*MOT:   right .
*CHI:
*MOT:   that's a donkey .
*CHI:   donkey
*MOT:   what's this [=duckie] ?
*CHI:   duckie
*MOT:   what does a duckie say ?
*CHI:   quack@o
```

(b) Model output

Figure 5: Further comparison of Childes data and the responses generated by the model under discussion

- Answer research question (answer probably affirmative, though not definitively).
- Maybe suggest multi-resolution exemplars augmented with video data using restricted boltzmann machines as possible alternative to explicit representations.

References

Rens Bod and Remko Scha. Data-oriented language processing: An overview. Technical Report LP-96-13, University of Amsterdam, 1996.

Nancy Chang and Olya Gurevich. Context-driven construction learning. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, chicago, 2004.

Hubert L. Dreyfus. *What Computers can't do: A Critique of Artificial Reason*. Harper and Row, 1972.

Jerry A. Fodor. *The Language of Thought*. Thomas A. Crowell Co., New York, 1975.

Jerry A. Fodor. Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioral and Brain Sciences*, 3:63–73, 1980.

R. Jackendoff. *Semantics and Cognition*. MIT Press, Cambridge/MA, 1983.

Manfred Krifka. Compositionality. In Robert A. Wilson and Frank Keil, editors, *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press, 1999.

George Lakoff and Mark Johnson. *Philosophy In The Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books, New York, 1999. ISBN 0465056733.

Brian MacWhinney and Catherine Snow. The cHILDES project: Tools for analyzing talk, 1995. URL `http://childes.psy.cmu.edu`.

Ernst Odolphi. Formal grammars for early child language, 2008. Bachelor's thesis, University of Amsterdam.

Richard Rorty. *Philosophy and the Mirror of Nature*. Princeton University Press, Princeton, 1979.

E. Rosch, C. B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439, 1976.

Luc Steels. Constructivist development of grounded construction grammar. In *Proceedings of the 42nd Annual Meeting of the ACL*, pages 9–16, 2004. URL `http://acl.ldc.upenn.edu/acl2004/main/pdf/159_pdf_2-col.pdf`.

Michael Tomasello. The item-based nature of children's early syntactic development. *Trends in cognitive sciences*, 4(4):156–163, April 2000. ISSN 1364-6613. URL `http://view.ncbi.nlm.nih.gov/pubmed/10740280`.

Michael Tomasello. *Constructing a Language : A Usage-Based Theory of Language Acquisition*. Harvard University Press, March 2005. ISBN 674017641. URL http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20\&path=ASIN/0674017641.

A. M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, October 1950. URL http://www.jstor.org/stable/2251299.

Andreas van Cranenburgh, Arjan Nusselder, Nadya Peek, and Carsten van Weelden. Towards a computational model for early language acquisition, 2007. URL https://unstable.nl/andreas/ai/2p/laac/verslag.pdf. 2nd year Bachelor of AI project, University of Amsterdam.

Jacqueline van Kampen. The learnability of syntactic categories. In *Proceedings of GALA*, pages 245–256, 2003.

Jacqueline van Kampen and Remko Scha. Modelling the steps of early syntax acquisition. In D. Cochran and R. Bod, editors, *Proceedings of the Workshop Exemplar-Based Models of Language*, Dublin, August 2007.

Jasper van Turnhout. Learning in dialog: a semantic grammar approach, 2007. Bachelor's thesis, University of Amsterdam.

Ludwig Wittgenstein. *Philosophical Investigations*. Macmillan, New York, 1953. G. Anscombe, translator.