UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Towards a Quatitative Corpus-based Evaluation Measure for Syntactic Theories

Permalink

https://escholarship.org/uc/item/9jm0f2cx

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN

1069-7977

Authors

Chang, Franklin
Lieven, Elena
Tomasello, Michael

Publication Date

2005

Peer reviewed

Towards a Quantitative Corpus-based Evaluation Measure for Syntactic Theories

Franklin Chang (chang@eva.mpg.de), Elena Lieven (lieven@eva.mpg.de),
and Michael Tomasello (tomas@eva.mpg.de)
Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6, Leipzig 04103, Germany

Abstract

It is difficult to quantitatively evaluate the match between syntactic theories and human utterance data. To address this problem, we propose an automatic evaluation measure based on a syntactic theory's ability to predict the order of words in actual utterances in corpora. To test this measure, a lexicalist theory of syntax acquisition is utilized and shown to achieve relatively high levels of prediction accuracy on both adult and child data from five typologically-different languages. Application of this evaluation measure to different input and output sets shows that this measure can address important linguistic criteria such as computational sufficiency, learnability, and typological-flexibility.

Keywords: Corpora, Syntax Acquisition, Evaluation

Comparison of Syntactic Theories

There is a fairly large gap between computational implementations of syntactic theories and human syntactic knowledge. For example, computational systems often use tagged corpora for either training or evaluation, while humans are able to demonstrate syntactic competence by processing or producing untagged utterances. Computational systems are tested on large written corpora or hand-crafted test sentences which may not be representative of the utterances that humans actually produce. Also children use their developing syntactic knowledge to produce utterances, but existing computational systems cannot be easily compared to these sometimes ungrammatical utterances (e.g., Anne, 2;7 "can you know doesn't be in there"). Finally, children can automatically learn any human language, while computational systems are often language-specific and are typically not tested with typologically different languages. To address these issues, we will present an algorithm that can automatically learn syntactic constraints from untagged corpora, and such knowledge can help in sentence production of utterances from 5 typologically different languages from both children and adults.

Syntax in sentence production helps to order the concepts that one wants to convey. For example, in Japanese, the verb must follow its arguments (e.g., "nobuko-ga inu-o suki" is ok, but not "nobuko-ga suki inu-o"). In German, articles must precede the nouns that which they agree in gender and which they mark the case of (e.g., "Die Tigerente isst den Döner" is fine, but not "Tigerente den isst Döner die"). In English, "the ball red" is ungrammatical, unlike the canonical "the red ball". Our evaluation measure will measure how well a syntactic learner can order the words in actual utterances in corpora. To insure that our test sentences are representative of the syntactic knowledge that is typically present in production, we will use the utterances in spoken corpora between children and adults (rather than artificial linguistic examples, or highly edited utterances in written corpora). By measuring the difference between the syntactic learner's predictions and the actual utterances, we can quantitatively evaluate the learner's viability as theory of syntax acquisition. To determine if this measure of syntactic knowledge is useful, we will use it to examine several criteria that are used to evaluate syntactic theories: computational sufficiency, typological flexibility, and learnability from the input.

One of Chomsky's (1957) early claims was that the computational sufficiency of a learner for representing human languages was an important determinant in selecting a language learner. If our computational learner is not appropriate for accounting for the utterances in real corpora, then we should expect that the learner will never achieve a high level of prediction accuracy. However, if it is possible to find a set of conditions where it is possible to predict most of the syntactically-appropriate word orders in a corpora, then such an algorithm could be said to be computationally sufficient for accounting for the syntactic constraints implicit in these orderings.

Another important criterion for a linguistic theory is that is can accommodate constraints that are embodied in the syntax of typologically-different languages (fixed/free word order, argument omission, rich/poor morphology). Because these typological features have an impact on the set of words and their order, our prediction measure can examine whether a particular learner can deal with the structural changes that these features induce. A typologically-flexible algorithm should yield similar magnitudes of improvements for languages with different typological features.

The third criterion relates to learnability or the relationship between the input and the syntactic abstractions that need to be learned. Traditionally, syntactic abstractions have been presumed to be difficult to learn because the input does not directly model the syntactic abstractions that are thought to be needed (this is called *the poverty of the stimulus* argument). Our measure can evaluate learnability by examining the prediction accuracy for the child's utterances when trained on the adult utterances from the same corpus. If the small sample of adult utterances in these corpora (small relative to the years of input that children use to learn to produce their utterances) can predict a substantial proportion of the child's utterances, then that would suggest that the input provides much of what the

learner needs to predict syntactic orderings, and that would militate against the view that the input is impoverished. In sum, word order prediction accuracy can evaluate how well a computational syntax learner can predict actual utterances, and by comparing different types of inputs, outputs, or languages, we can use this measure to examine issues related to computational sufficiency, learnability, and typological-flexibility.

To demonstrate the viability of our evaluation measure, we will test a simple computational theory of syntax acquisition that is inspired by connectionist approaches to syntax (Elman, 1990). To avoid the complications of learning and using abstract syntax, this approach will only represent the lexically-specific aspects of these theories. Lexically-specific representations are compatible with most linguistic approaches and are easily collected from language corpora. But while there are good motivations for taking this approach, there are many reasons to think it will not be very successful at predicting syntactically-appropriate word orders. While many linguistic theories emphasize the role of lexical entries in syntax (Pollard & Sag, 1994), these lexical entries get their power from links to more abstract information (e.g. verb subcategorization, parameters). And while there is evidence showing that children make extensive use of lexical information early on in language acquisition (Bates & Goodman, 1997; Tomasello, 1992), it is also thought that abstract syntactic structures/constructions and meaning, are also needed for acquisition of the appropriate constraints (Gleitman, 1990; Lieven, Behrens, Speares, & Tomasello, 2003; Pinker, 1989; Tomasello, 2003). Also, distributional corpus-based approaches that attempt to discover syntactic categories like nouns and verbs have not yielded techniques that are comparable across languages (part of speech taggers often use different sets of tag categories for each language, Manning & Schütze, 1999) and approaches that attempt to be psycholinguistically realistic have not been shown to yield the same level of accuracy in typologically-different languages (Mintz, 2003; Redington, Chater, & Finch, 1998).

Computational distributional learning approaches have been mainly tested in English, a language that is ideal for these algorithms. English has relatively rigid word order, limited morphology, a small set of articles, and no argument omission. These properties make the previous word a particularly good cue for classifying words into categories like nouns and verbs in English. But it is not clear whether these techniques will work for other languages. In relatively free word-order languages like Japanese, German, or Croatian, there might be more variability in these word transitions. In languages with rich morphology like Croatian, nouns can have different forms for each combination of case, gender, and number, and that means that a larger number of word transitions have to be abstracted over in order to discover a category. And in languages that allow all arguments to be omitted like Japanese (relatively free word-order) and Cantonese (relatively fixed word-order), the words that appear before the word to be categorized will be more variable, and this

makes it harder to use the previous word to create a category. Because of these issues with distributional approaches to syntax, it is not clear that there is an automatic way to learn syntactic knowledge cross-linguistically from distributional information in corpora.

To summarize, we have proposed an evaluation measure that should allow us to examine various linguistic criteria that are important in comparing syntactic theories. But because existing computational instantiations of syntactic theories do not do word order prediction, we will present a simple example of the syntax acquisition learner that can accomplish the mapping that we are suggesting that children are learning. This learner, which we called the lexical producer, will be shown to predict the order of words in the speech of adults and children in a set of typologically-different languages. Then we will show that some of these constraints can be learned from just a small subset of the adult speech that these children actually receive. This will demonstrate that it is possible to create a system that achieves reasonable performance when evaluated by this measure. In the future, when existing generative and statistical syntactic theories are adapted to word order prediction, this evaluation measure will allows us to see how much different syntactic abstractions (e.g., parameters, syntactic categories) improve prediction over the lexical producer.

The lexical producer algorithm

Since our goal is to model the acquisition of the constraints needed for ordering words in sentence production, our approach needs to fit with work in sentence production (Bock, 1995). First, sentence production begin with a message or meaning that the person wants to convey. This message is important in determining the words that are used. Second, production is an incremental process where structure selection is determined by the words that are selected early on in an utterance. A third feature of sentence production is that we need to "deactivate the past" (Dell, Burger, & Svec, 1997), to keep recently produced (and therefore activated) words from being reactivated. To incorporate these aspects of production into the lexical producer, we formalized the problem in this way. The lexical producer was given the words that the speaker actually produced, but as an unordered list that we called the *candidate set*. The words that a speaker uses reflects features of the message that they are trying to convey (e.g. articles encode discourse information), and since speakers know what they want to say before they speak, the candidate set captures the influence of message knowledge. The task of the producer is to create an ordered sequence of words from this set. This ordering will be done incrementally, where the system has to choose the next word in the utterance from the candidate set. Then the word that was actually produced next by the speaker will be removed from the candidate set, and then the system will again attempt to choose the next word from this shorter set. The removal of this word accomplishes the suppression of the past, and

allows us to test the system's accuracy against each of the choices that the speaker actually made. This is crucial for evaluating the system against actual utterances in different languages.

Before describing the algorithm in more detail, it is first important to mention two issues related to the preprocessing of the utterances. Before processing each utterance (both for production and for recording the statistics), the punctuation (period, question mark, exclamation mark) was moved from the end of the utterance to the first position. This made it the first word in the utterance, and the system could use this punctuation word to predict the first real word in the utterance (e.g. "? who did this"). Second, repeated words within an utterance were given number indexes (e.g. the-1) to distinguish them from other tokens, but these indexes started from the first repeated word starting from the end of the utterance (e.g., ". i-'m the-1 king of the castle"). This made production of the latter part of an utterance (e.g. "of the castle") similar to other utterances with those words.

The lexical producer algorithm was inspired by a particular connectionist architecture for sentence production. This architecture, called the Dual-path architecture, has been argued to be superior for sentence production models, because it allows these models to generate utterances with words in novel sentence positions (Chang, 2002). This architecture is composed of sequencing pathway, which is a simple recurrent network that learns which sets of words occur in different sentence positions (Elman, 1990), combined with a message pathway that activates message-appropriate words. By combining the output of these two pathways, the system could use novel meanings to sequence novel utterances, but in a way that was consistent with the syntactic properties of the language. To simulate the two pathways in the Dual-path model, two types of information will be used by the lexical producer to sequence the words in the candidate set. One type of information, referred to here as *context* information, will record how well the context (the previous word) predicts the next word in the sentence, much like the sequencing network in the Dual-path model. The other information, referred to here as *access* information, will record how often a word precedes other words in an utterance. When the access information for all the candidate words is combined together, it simulates the competition between the words that are activated by the message in the message pathway in the Dual-path model.

The lexical producer uses only lexically-specific information in the input corpora to derives the context and access information. To do this, it collects three types of lexical statistics from the corpora: the *context counters*, the *access counters, and the paired counters*. The context counters count how often a pair of adjacent words in an utterance occur together in utterances in the corpora. Taking as an example the utterance ". look at me", the context counters for .-look, look-at, and at-me pairs would all be 1. The access counter incremented its counter for all word pairs (excluding the punctuation word) where the first

member A was before the second member B (A>B). So in the above utterance, the access counters for look>at, look>me, or at>me pairs would all be 1. The paired counter recorded how often each pair of words in the utterance occurred together in any order. The paired counters for .=look, .=at, .=me, look=., look=at, look=me, at=., at=look, at=me, me=., me=look, and me=at pairs would all be 1. These three types of counters were used to create the context and access biases that will be used by the lexical producer. The context bias for an ordered pair of words was just the context counter divided by the paired counter for those two words (e.g. context-bias(look-me) = 0/1 = 0). The access bias for an ordered pair of words was just the access counter divided by the paired counter for these two words (e.g. access-bias(look>me) = 1/1 = 1). Dividing by the paired counter removed the pair frequency from the biases.

Now the production algorithm will be described. Before production began, the producer started with the candidate set (e.g. "at", "me", "look") and the punctuation word (e.g. ".") as the previous word. For each word in the candidate set, the system calculated a context score and access score and then summed them together to get a choice score. The context score for each word in the candidate set was just the context bias from the previous word at this point in production of this utterance. The access score for each word in the candidate set was just the sum of all the access biases to that word from all the words in the candidate set divided by the number of candidate words. Hence, the context and access scores captured different aspects of ordering. The context score ranked the candidate words in terms of how likely they were to occur after the previous word. The access score ranked the candidates in terms of how likely they were produced before the other candidates.

The context bias is a type of bigram statistic that is commonly used in distributional learning approaches to syntax. The access statistic on the other hand is relatively novel, because it assumes that the existence of a candidate set of words that are in competition for selection. Since most distributional learning systems that use corpora take a comprehension approach to syntax (utterance -> abstract representation), they have not made use of access-type statistics.

The corpora

To allow us to compare typologically-different languages, corpora for English, German, Croatian, Japanese, and Cantonese were selected from the CHILDES database (MacWhinney, 2000). Our goal for this comparison was just to see if the algorithm could work at a similar level of accuracy for languages that should be more difficult for distributional approaches than English. The databases that were chosen were those that attempted to gather substantial amounts of recorded data from a single child. Table 1 shows the children from the different databases that were chosen as the corpora to be used in our analysis. The utterances in each database were separated into child (utterances from the target child) and adult (all other

utterances in the corpora). The final column in Table 1 specify the number of child utterances that were used. The utterances in each database had extra coding of missing sounds, pause information, and a variety of other codes. Utterances which had uncodeable parts were excluded. Other codes were stripped so that the original utterance as the child and adult said it was preserved. Morphologically marked words were left unchanged (e.g. boy-s), and in general all space-separated strings were treated as different words. All words were converted into lowercase.

Table 1. Corpora from individual children in 5 languages.

Corpora/ Child	Database	Age	# of Child Utt.
English Anne	Manchester (Theakston, Lieven, Pine, & Rowland, 2001)	1;10-2;9	19943
English-Dense Brian	MPI-EVA (Lieven et al., 2003)	2;0-3;11	174110
German Simone	Nijmegen (Miller, 1976)	1;9-4;0	28561
German-Dense Leo	MPI-EVA	1;11-4;11	139540
Croatian Vjeran	Kovacevic (Kovacevic, 2003)	0;10-3;2	20875
Japanese Ryo	Miyata (Miyata, 1992, 1995)	1;4-3;0	11778
Cantonese Jenny	CanCorp (Lee, Wong, Leung, Man, Cheung, Szeto, & Wong, 1996)	2;8-3;8	10021

The utterances from all seven corpora were used to train and test individual versions of the model. Utterance accuracy was the dependent measure used in all the analyses and it was the percentage of utterances correctly predicted (where each word in an utterance had to be correct for the utterance to be correct). The same word with different number indexes were treated the same for accuracy (e.g. the-1 = the). Since candidate sets with only one word are trivial to produce and inflate the accuracy results, these situations (both one-word utterances and the last position of multi-word utterances) were excluded from the results. To get measure of chance for each of these corpora, a Chance model was created that made a random word choice from the candidate set at each position in each utterance in each corpus. For example, if you have only two words in your candidate set, then you have a 50% chance to get them in the right order.

The first step in evaluating the lexical producer is to see if it is computationally sufficient for predicting word order. As mentioned earlier, it is not clear that it is possible to find a single set of lexically-specific statistics that can predict the order of words in actual utterances in different languages. To address this issue, we will test whether the adult's or the child's data can be used to predict itself. To do this for both the adult and the child, we collect our two lexical statistics by passing through the data, and then we test the system with the same data. Self-prediction tests how well the learner can memorize and reproduce the test set, and is akin to repetition tasks that have been used to test syntactic knowledge in children and adults (Chang, Bock, & Goldberg, 2003). Because it provides all the words and the orders that are needed in the test utterances, self-prediction should tell us whether the lexical producer is sufficient for the task of predicting word order in raw corpora.
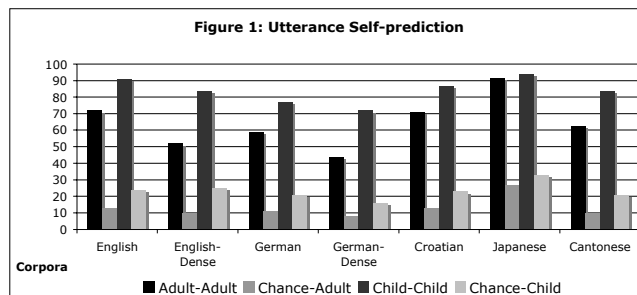

Figure 1: Utterance Self-prediction

Figure 1 shows the utterance accuracy during self-prediction. Overall, the algorithm improves utterance prediction over chance by 43% for adults and 58% in the children. What these results show is that both children and adults tend to be fairly consistent in the order of words that they use. For example, the English-Dense child said "Bill and Sam", but never "Sam and Bill". If the child had used both grammatical orders for a large proportion of their utterances, their utterance accuracy would be closer to 50%. Instead, the lexical producer can predict 81% of the child utterances (Child-Child) with a single order and 53% of the adult utterances (Adult-Adult). The fact that the children were more likely to use one order for any set of words suggests that their representations are more lexically-specific than the adults around them. The adult utterances are more order variable, presumably because the adults have a more abstract syntax which allows them to produce differ word orders to convey different meanings. But overall, these results show that the two lexically-specific statistics in the lexical producer are sufficient to account for most of the syntactically-appropriate word orderings in both adults and the children.

The next question is whether the lexical producer system has the right learnability characteristics for syntax acquisition from sparse amounts of adult data from real corpora. To test this, the same algorithm was applied to the adult data to extract statistics that were used to predict the child's utterances. It would be surprising if this were possible, since children and adults use different words and talk about things in different ways. Furthermore, our samples of adult speech are only a small sample of the input the child actually receives (Cameron-Faulkner, Lieven, & Tomasello, 2003).

However when we test the lexical producer on the child's data when trained on the adult input (Adult-Child), we find that it can use the adult data to increase utterance accuracy 36% over chance (Figure 2). While it is not surprising that the adult input is useful to predict the child's output (since

the child is learning the language and the words from these adults), it surprising that more than half of the distance between chance (Chance-Child) and self-prediction (Child-Child, same as Figure 1) can be predicted from a small sample of adult speech to the child without abstract categories or structures like syntactic trees, constructions, meaning, or discourse. Surprisingly, the amount of input does not seem to change performance, as the dense corpora have the same accuracy levels as the non-dense versions. Even though the dense corpora have more utterances to learn from, they also have more utterances to account for, and therefore the percentage accounted for does not change much.

Figure 2: Utterance Accuracy on Child's Data depending on Input
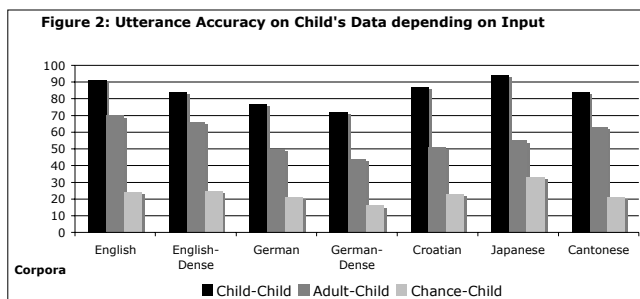
To examine typological differences between languages, multiple corpora from each language will need to tested and statistically compared. But within the seven corpora tested, there does not seem to be a large variation in the prediction accuracy of the lexical producer across the different languages. One reason for this is that that evaluation measure and the lexical producer balances out some of the typological differences. In argument-omission languages, one has less distributional information, but one has also fewer word orders to predict. In languages with rich morphology, the context and access statistics in these languages should also be more specific and reliable, but any particular pairing of words is less likely to be represented in the input corpus.

To summarize, the lexical producer is able to account for, averaging over all the corpora, 58% of the child's utterances that are two words or longer when trained on a small subset of the adult data that the actual child receives. If we include all the one-word utterances, the lexical producer's accuracy rises to 76%, which amounts to 309,559 correctly produced utterances in five typologically-different languages. By themselves, these results may not seem impressive. But if we realize that all existing linguistic theories assume that word order would require some abstraction over words, it is surprising that more than three quarters of the utterances produced by children in five languages are predictable from a small sample of adult data with no abstractions at all.

Conclusion

Theory evaluation in linguistics is not a quantitative science. To have a quantitative science, one needs a way to numerically evaluate the fit between data and theories. Here, we have provided an evaluation measure of the fit

between the order of words in utterances and computational learner of syntax that allowed us to examine linguistic criteria related to computational sufficiency, learnability, and typological-flexibility. The measure was high when the training and test data were the same, low when the system was randomly choosing the order of words, and intermediate when trained on adult data to predict the child's utterances. And the measure was able to achieve high levels for the five typologically-diverse languages suggesting that it can be used to compare the learning of different languages.

In addition, we introduce a simple lexicalist theory (the lexical producer) and showed that it was able to achieve a good fit to the data. The lexical producer was inspired by the way that the psycholinguistically-motivated connectionist model of sentence production, the Dual-path model (Chang, 2002), learned different types of information in each of its pathways. The lexical producer during self-prediction was able to predict a large percentage of word order patterns, which suggests that lexically-specific knowledge can do much of the work of predicting the order of words in human language. When trained on adult speech to the child, the algorithm was able to predict more than half of the utterances that the child produced. This suggests that the constraints that govern the child's utterances are learnable from the meager input that this algorithm was given. And finally when tested on languages which might be difficult for distributional learning (flexible word order languages like German, or morphologically rich languages like Croatian, or argument-omitting languages like Japanese), the lexical producer was able to acquire word order constraints at a level that approximates its behavior in English. This suggests that it is a typologically-flexible algorithm for syntax acquisition.

There are three respects in which this work is novel. First, it shows that in the utterances from children in five typologically-different languages, there is a tendency to use a single word order for any set of words. Previous work in English suggests that children tend to be conservative and lexically-specific in their sentence production (Lieven, Pine, & Baldwin, 1997; Tomasello, 1992), but this is the first typologically-diverse demonstration. Second, it shows that a small sample of the adult-speech to children can predict more than half of the child's utterance orderings. This is important cross-linguistic counterevidence to the claim that there is not enough input to explain syntax acquisition without innate knowledge (poverty of the stimulus). The third result is that lexical access can play an important role in computational approaches to distributional learning. Lexical access is often thought to be a performance component of sentence production, rather than a part of syntax acquisition. But since speakers must learn to order their ideas (and the words that they activate) and order of words in the input is useful for learning these constraints, it suggests that lexical access knowledge should play a part in theories of syntax acquisition and use.

The lexical producer only represented the lexical aspects of the Dual-path model, and hence it is not a complete

syntactic theory. In particular, it is not sufficient for generating utterances where meaning is used to generate a novel surface form. Its success has more to do with the conservative nature of language that people actually use, rather than its learning mechanism. That is why we think that the important aspect of this work is the evaluation measure, which allows us to measure syntactic knowledge in untagged utterances by children or adults in different languages. Hopefully, this evaluation measure will encourage advocates of particular syntactic theories to show that their representations can be learned from untagged input in different languages, and that these representations can be use to order the words in sentence production.

Acknowlegements

References

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language & Cognitive Processes, 12*(5-6), 507-584.

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 181-216). Orlando, FL: Academic Press.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27*(6), 843-873.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science, 26*(5), 609-651.

Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition, 90*(1), 29-49.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104*(1), 123-147.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179-211.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition: A Journal of Developmental Linguistics, 1*(1), 3-55.

Kovacevic, M. (2003). Acquisition of Croatian in Crosslinguistic Perspective. Zagreb.

Lee, T. H. T., Wong, C. H., Leung, S., Man, P., Cheung, A., Szeto, K., et al. (1996). *The development of grammatical competence in Cantonese-speaking children*. Hong Kong: Dept. of English, Chinese University of Hong Kong. (Report of a project funded by RGC earmarked grant,1991-1994).

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*(2), 333-367.

Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24*(1), 187-219.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk, Vol 2: The database* (3rd ed. ed.).

Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press.

Miller, M. (1976). *Zur Logik der frühkindlichen Sprachentwicklung: Empirische Untersuchungen und Theoriediskussion*. Stuttgart: Klett.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117.

Miyata, S. (1992). Wh-questions of the third kind: The strange use of wa-questions in Japanese children. *Bulletin of Aichi Shukutoku Junior College, 31*, 151-155.

Miyata, S. (1995). The Aki corpus -- Longitudinal speech data of a Japanese boy aged 1.6-2.12. *Bulletin of Aichi Shukutoku Junior College, 34*, 183-191.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425-469.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28*(1), 127-152.

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

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.