524

Johansson, B. S. 1977. Levels of mastery of the coordinators 'and' and 'or' and logical performance test. *British Journal of Psychology*, 68, 311-320.

Noveck, I. A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2), 165-188.

Papafragou, A. 2003. Aspectuality and scalar structure. *Proceedings of the Annual Boston University Conference on Language Development*, 27(2), 626-637.

Rooth, M. 1992. A theory of focus interpretation. *Natural Language Semantics*.1, 75-116.

Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, 30, 191-205.

Villalta, E. 1994. Plural indefinites in Spanish and distributivity. Ms. University of Paris 7.

A Dynamic Learning Model for Categorizing Words Using Frames

Hao Wang and Toben H. Mintz
University of Southern California

1. Introduction

Grammatical categories (e.g., noun, verb, and etc) are the building blocks of grammar, as grammars organize abstract linguistic categories rather than individual lexical items. Learning the categories of words is, thus, fundamental to the acquisition of language. Even theories in which learners are endowed with considerable innate linguistic knowledge allow that children need information from the input to map lexical items to grammatical categories. While a number of sources of category information have been proposed (including phonological information, morphological information, and semantic information), one information source that has received considerable attention recently is distributional information from lexical co-occurrence patterns (Chemla, Mintz, Bernal, & Christophe, in press; Redington, Chater, & Finch, 1998, Mintz, 2003; Mintz, Newport, & Bever, 2002). As an example, consider the sentence in (1).

(1)   The cat is on the mat.

A distributional learner might notice that the word *cat* and *mat* appeared in the same context—immediately following the word *the*. Such a learner would categorize *cat* and *mat* together, based on the overlap in distributional context, and would also place all words immediately following *the* throughout a corpus in the same category; the resulting category would contain almost exclusively nouns.

The idea that grammatical categories are related to distributional patterns is not new. Modern versions of this idea go back to structural linguistics in the early and mid-20th century (Harris, 1951), and an explicit formulation of the role of distributional information in language learning was put forth by Maratsos & Chalkley (1980). Over the past decade, increases in computer power and the availability of digital databases of speech input to children (e.g., MacWhinney, 2000) have made it possible to test how well a variety of distributional patterns could serve as a bases for categorizing words early in language development.

For example, researchers have investigated how contexts preceding a target word differ in informativeness from contexts that follow, how proximal contexts differ from distal contexts, whether considering more contexts is beneficial over considering just a few, etc (Cartwright & Brent, 1997; Chemla et al., in press; Redington et al., 1998; Mintz et al., 2002; Mintz, 2003). These investigations have helped to delineate the kinds of patterns in actual child-directed speech that could serve as a basis for categorizing words.

A particularly robust distributional pattern is the *frequent frame* (Mintz, 2003). Mintz (2003) defined a frame as two jointly occurring words with one word intervening, and a frame was *frequent* in a given corpus if it surpassed a threshold frequency in that corpus. Mintz proposed that learners could categorize together target words that occur within the same frequent frame throughout a corpus. For example, (2) shows several utterances from the Eve corpus (Brown, 1973) in the CHILDES database (MacWhinney, 2000) that contain the *you_X_the* frame—a frequent frame in that corpus—and the words in the X position are all verbs. Mintz analyzed six corpora from the CHILDES database, and found that frequent frames form categories in which members almost exclusively belong to the same linguistic category. For example, (3) shows the words categorized by the *you_x_it* frame in the Peter corpus (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbown, & Hood, 1975). The category contained 433 word tokens ranging over 93 types, all of which were verbs. Not all frame-based categories were perfect, but their accuracy was very high across all corpora.

(2)   *you_x_the*

Would you put the cans back?
You get the nuts.
You take the chair back.
You read the story to Mommy.

(3)   *you_x_it*

put, see, do, did, want, fix, turned, get, got, turn, throw, closed, think, leave, take, open, find, bring, took, like, knocked, putting, pull, found, make, have, fixed, finish, try, swallow, opened, need, move, hold, give, fixing, drive, close, catch, threw, taking, screw, say, ride, pushing, hit, hiding, had, eat, carry, build, brought, write, wiping, wipe, wind, unzipped, underneath, turning, touching, tore, tie, tear, swallowed, squeeze, showing, show, said, rip, read, reach, pushed, push, play, pick, parking, made, love, left, knock, knew, hid, flush, finished, expected, dropped, drop, draw, covered, closing, call, broke, blow

The success of frequent frames in categorizing words makes the pattern a prime candidate for a distributional source of category information in language acquisition. Nevertheless, the frequent frame analysis procedure proposed by Mintz (2003) was not intended as a model of acquisition, but rather as a demonstration of the information contained in frequent frames in child-directed speech. Its limitations as a psychologically plausible model include the fact that it requires two passes through a corpus, once to identify the frequent frames by tallying frame frequency, then again to categorize words. In addition, in the first pass it tracks the frequencies of all the frame types. Since a given corpus contains a large number of different frames—e.g., there are approximately 21,000 frame types in the Peter corpus—this is probably not a realistic computation for children. Hence, Mintz (2003) did not address the question of whether an actual learner could detect and use frequent frames to categorize words.

Some behavioral evidence suggests that infants can detect and use frames. For instance, Gómez and Maye (2005) show that 15 month-olds are sensitive to frame-like units in highly controlled experimental material, and frames have been shown to facilitate categorization in adults and infants (Mintz, 2002, 2006, 2007). But none of these studies directly addressed the issue of detecting frequent frames in children's normal input. This paper addresses this question with the investigation of a computational model of frequent frame detection that incorporates more psychologically plausible assumptions about the memorial resources of learners. In addition, it implements learning as a dynamic process that takes place utterance by utterance as a corpus is processed, rather than "in batch" over an entire corpus. The results suggest that the frequent frame pattern is a robust categorization context even under these more severe processing constraints.

2.   The Model

In this paper, we propose a more plausible frame-based model of word categorization by taking the following facts into account. First, children possess limited memory and cognitive capacity and cannot track the occurrences of all the frames in a corpus. Second, memory retention is not perfect: frames represented in memory may be forgotten if they are encountered infrequently. We implement these constraints by limiting to 150 the number of different frame types (and their frequencies) that can be held in memory at one time. We also implement a forgetting function such that, all else being equal, frames in memory that have not been encountered recently are less likely to stay in memory than frames that have recently been encountered. A necessary corollary to these constraints is that learning is dynamic and incremental, and is influences by the ordering of utterances in the corpus, as opposed to a batch process that considers the entire corpus as a whole.

We present two analyses of the application of the model to a set of corpora. The computational procedure is identical for both analyses and is presented first.

2.1. Procedure

Child-directed utterances from each corpus were processed individually. Utterances were presented to the model in the order of appearance in the corpus, which is along the age of the child. First, each utterance was segmented into frames. For example, (4) shows the segmentation of an utterance into frames

If the memory is not full, a newly-encountered frame is added to the memory and its initial activation is set to 1. If an encountered frame already exists in the memory, its activation increases by 1. To simulate a "forgetting" function, at each processing step, the activation of all frames in memory decreases by .0075.

(4)   You read the story to mommy.
      You   X   the
            read X   story
                the X   to
                    story X mommy

Since the memory buffer only stores the most active 150 frames, it becomes full very quickly—after processing about 50 utterances. When the memory is full, a newly-encountered frame replaces the least active frame with activation less than 1. If all activations are greater than 1 in the memory, no changes take place other than the usual decrease in activation of all frames. Table 1 shows a sample memory buffer.

Table 1 A example of memory buffer

Activation	Frame
8.9992	the_X_to
6.9984	read_X_story
5.9976	you_X_the
0.7968	story_X_Mommy
...	...

2.2. Input corpora

In order to compare our model's performance to the results of Mintz (2003), we used the same six corpora from the CHILDES database (MacWhinney, 2000) analyzed in Mintz (2003): Eve (Brown, 1973), Peter (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbown, & Hood, 1975), Naomi (Sachs, 1983), Nina (Suppes, 1974), Anne (Theakston, Lieven, Pine, & Rowland, 2001), and Aran (Theakston et al., 2001). The input to the model is adult speeches in each

corpus when the child was 2;6 or younger. This is the period when children start to show some knowledge of categories in their production.

3.   Analysis 1: Evaluation of categorization

The model's performance was evaluated after processing every 100 frames and at the end of the processing. To evaluate how accurate the categorization is, each word was labeled with its actual linguistic categories in the context. We used the labeling from Mintz (2003) where possible. For other words, a NLP tagger was used (Toutanova and Manning, 2000; Toutanova, Klein, Manning, and Singer, 2003). The actual linguistic categories we used are listed in Table 2.

Table 2 Grammatical categories used in the evaluation of categorization

Noun, Pronoun	Verb, Aux., Copula	Adjective
Preposition	Adverb	Particle
Cardinal number	Determiner ·	Wh-word
Negation ("not")	Conjunction	Interjection
Foreign word		

Next, the accuracy of categorization was computed. Accuracy is a standard evaluation metric and was used in Mintz (2003) and elsewhere. Accuracy was computed with the equation in (5). In each frame-based category, every pair of word tokens categorized by the memorized frames was compared. Each pair was classified as a *hit* if the two words were from the same grammatical category, or a *false alarm* if not. Basically, accuracy is penalized when two words from different grammatical categories are grouped together. The maximum value of accuracy is 1. High accuracy indicates that most of the words appeared in a frame are belong to the same grammatical category.

$$(5) \qquad \text{Accuracy} = \frac{\text{hits}}{\text{hits} + \text{false alarms}}$$

3.1. Results

Figure 1 shows the accuracies of the 6 corpora at the end of the processing, which are in the range from 0.93 to 0.97. The average accuracy is 0.96. These accuracies are quite high and are comparable to the results of Mintz (2003). It indicates that the model has achieved very good categorization after processing the entire corpus.
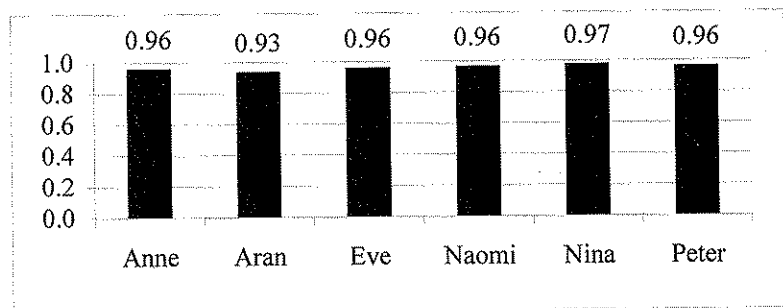
Figure 1 Accuracies after processing the six corpora

The most important feature of this model is that it processes the utterances dynamically. So we can see the development of accuracy along the processing. Figure 2 is the development of accuracy of Eve. It's a typical curve of accuracy development for the six corpora. The x-axis is how many frames have been processed. The y-axis is the accuracy. We can see that the accuracy is very high, around .96 in most part, and very stable in the entire process. Figure 3 shows the development of accuracy of Peter. It's similar to that of Eve, very high and stable in the entire process.
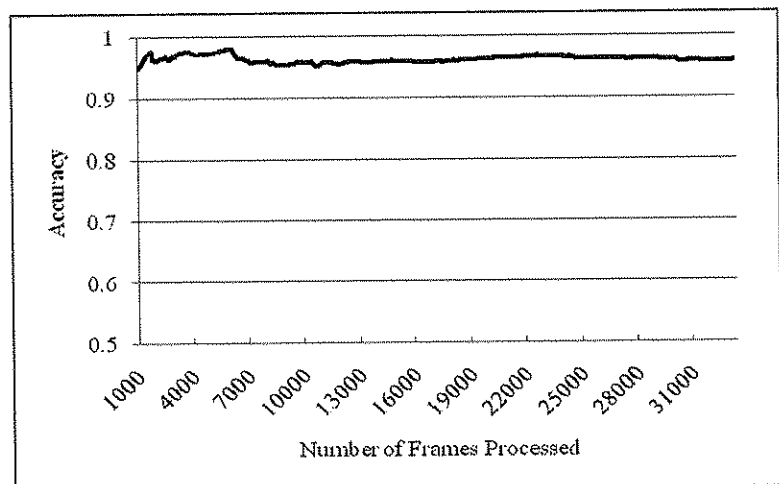

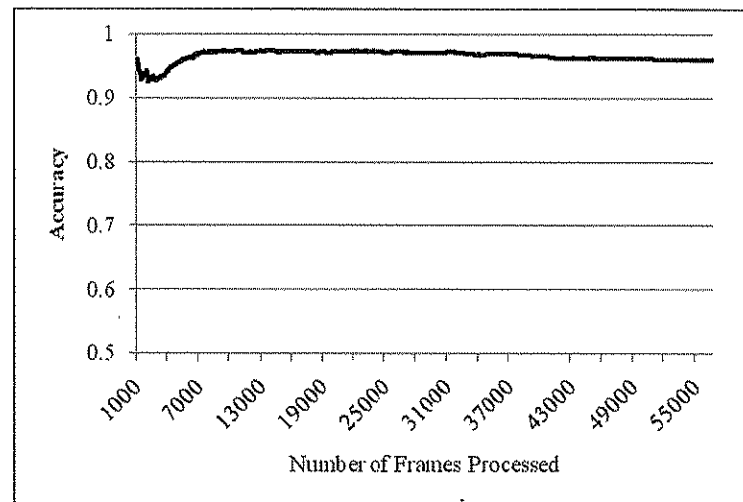**Figure 2 The development of accuracy of Eve**

Figure 3 The development of accuracy of Peter

4. Analysis 2: Comparison to frequent frames

In addition to looking at accuracy, we also compared the active frames in the memory to frequent frames (Mintz, 2003). The data and processing of the model are the same as Anal 1. After processing every 100 frames, all the frames in the memory were labeled with either frequent frame or not. We used the criterion from one analysis in Mintz (2003) that treated the 45 most frequent frames as the set of frequent frames. We then counted how many frames in the memory were frequent frames.

4.1. Results

In Figure 4 and Figure 5, the bold line is the percentage of frequent frames that appeared in most active 45 frames. The regular line is the percentage of frequent frames that appeared in the entire memory.

Figure 4 shows that after processing one third of Eve corpus, about 70% of frequent frames are in the most active 45 frames, and more than 80% of frequent frames are in the entire memory. At the end of the processing, approximately 90% of frequent frames are in the most active 45 frames, and there is only one frequent frame not in the memory. Figure 5 show similar results for the Peter corpus.
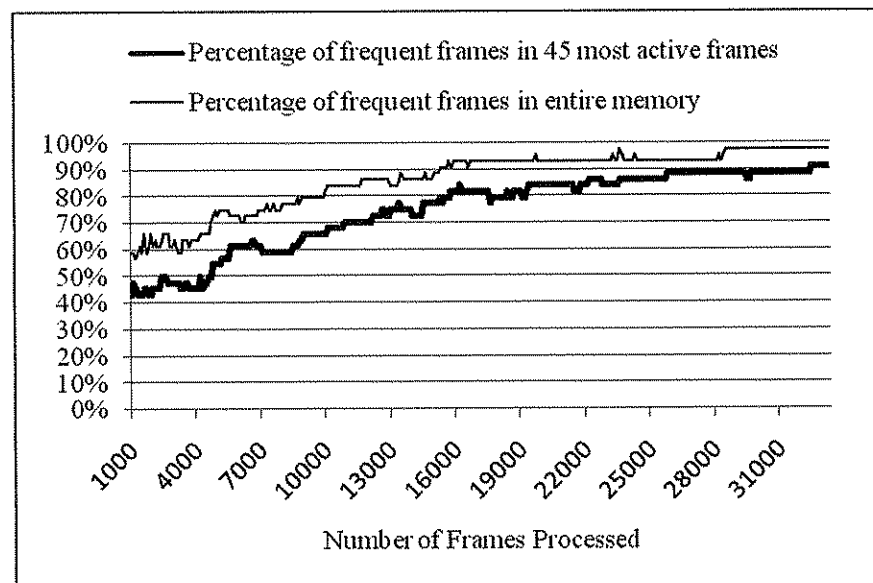
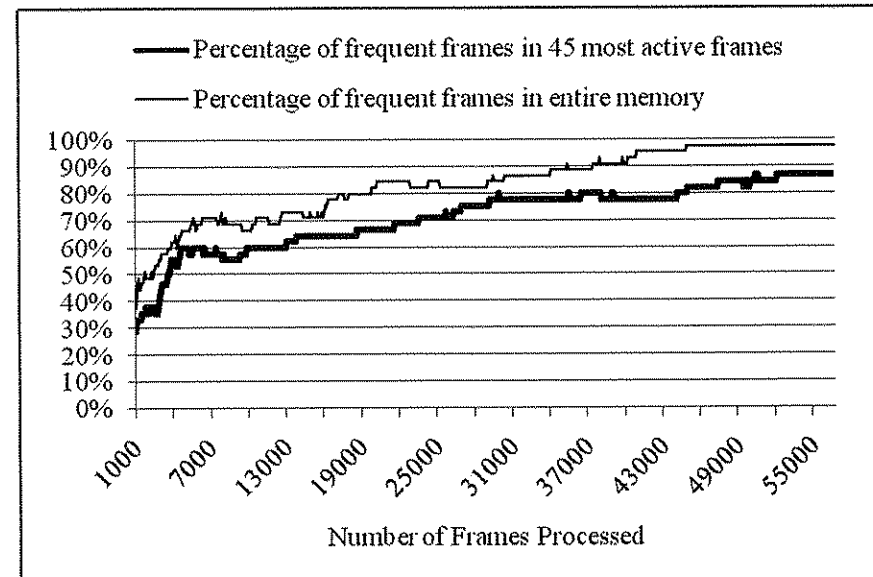**Figure 4 Percentage of Frequent Frames of Eve**

Figure 5 Percentage of Frequent Frames of Peter

Table 3 shows the ten most active frames of Nina at the end of processing. The asterisk beside a frame means that that frame is a frequent frame in Mintz

(2003). In fact, all the ten most active frames are frequent frames. A large number of word types or word tokens appeared in these frames. Hence they have very high frequency in the corpus. It is not surprising, therefore, they are the most active frames in the memory. It is also not surprising, then, that the resulting categorization accuracy is high, since the a large number of the active frames are frequent frames, which are known to be accurate categorizers.

We have shown that many frequent frames are in the memory. But what about frames in the memory not frequent frames? Table 4 shows the ten most active frames that are not frequent frames in that corpus. (Some frames that are not frequent frames in one corpus are frequent frames in other corpus. For example, the *put_x_in* frame is not a frequent frame in Nina, but it is a frequent frame in the 5 other corpora.)

The high accuracy achieved by our model suggests that these frames are also good categorization contexts. This is confirmed by visual inspection of the resulting categories. For example, Table 5 lists all the words appeared in the *we_x_the* frame, which is not frequent frame in any of the six corpora. All the contained words are verbs, and there are a relative large number of word tokens and word types. So active frames that do not meet a criterion for being a frequent frame can also be good categorization contexts.

Table 3 Ten most active frames of Nina at the end of processing

Activation	Frame	# of types	# of tokens
698.9	you_X_to*	33	735
520.9	what_X_you*	9	557
305.0	do_X_want*	2	341
251.4	you_X_the*	75	287
171.9	to_X_the*	53	207
130.0	you_X_me*	17	162
128.5	are_X_doing*	2	164
116.2	want_X_to*	11	149
115.9	you_X_it*	59	151
115.1	are_X_going*	3	138
108.7	what_X_is*	11	145

* means that frame is a frequent frame in Mintz (2003).

Table 4 Ten most active frames that are not frequent frames (Mintz, 2003) in that corpus

Nina	Eve	Peter
who_X_you	no_X_not	you_X_ta
you_X_what**	a_X_bit	ta_X_it
what_X_these	it_X_you	do_X_have**
what_X_we**	what_X_of	it_X_you
a_X_on**	where_X_you**	the_X_and**
we_X_the	are_X_gonna**	the_X_room
the_X_to**	on_X_floor**	do_X_see
put_X_in**	you_X_some**	on_X_floor**
I_X_the	is_X_a**	what're_X_gonna
what_X_to	out_X_the	I'll_X_it**

** means this frame is a frequent frame in other corpora.

Table 5 Intervening word and number of tokens of _we_x_the_ frame

put	36	set	1	fed	1
make	4	fill	1	see	1
bring	2	hang	1	attach	1
fix	2	comb	1	take	1
building	1	build	1	saw	1
have	1	chased	1		
give	1	open	1		

5. General discussion and summary

The analysis of frequent frames (Mintz, 2003) established that a certain type of distributional context, when it occurs frequently, provides informative category information. However, detection of those contexts required memory resources that are not feasible for actual child learners. The model we proposed here is more realistic in terms of its demands on learners, and it achieves comparable performance. The present model is also driven by frequency, but in a dynamical way. In this model, no frame can stay in the memory if it is not encountered relatively frequently. This explains why so many frequent frames occurred as active frames by the time the entire corpus was processed. However, using this dynamic processing approach we were able to discover informative frames which did not meet the criterion for frequent frames.

Given that frequency plays a role in both the "batch" and dynamic categorization methods, a broader question is why frequency should matter at all in detecting informative contexts (in this case, frames). One trivial answer is that useful categorization can only occur if a significant number of words are brought together in a category; under a system where categories are defined by contexts, this can only occur if the contexts in question occur frequently enough to cover a significant number of words. But a deeper answer is that frequently occurring frames indicate syntactically stable patterns: A frame is defined by a non-adjacent dependency, thus the frequent co-occurrence of two framing elements is likely due to an underlying structural cause, rather than by chance, and that underlying structure is also likely to constrain the intervening word. Words that appear in these frames are therefore are syntactically more constrained. See Chemla et al. (in press) for a more detailed discussion.

In summary, our model demonstrates very effective categorization of words. Even with limited and imperfect memory, the learning algorithm can identify highly informative contexts after processing a relatively small number of utterances, thus yield a high accuracy of word categorization. It also provides evidence that frames are a robust cue for categorizing words.

References

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (in press). Categorizing words using "frequent frames": What cross-linguistic analyses reveal about distributional acquisition strategies. _Developmental Science_.

Bloom, Lois, Hood, Lois, & Lightbown, Patsy M. (1974). Imitation in language development: if, when and why. Cognitive Development, 6, 380-420.

Bloom, Lois, Lightbown, Patsy M., & Hood, Lois. (1975). Structure and variation in child language. Monographs of the Society for Research in Child Development, 40(2), 1-97.

Brown, Roger W. (1973). A first language: The early stages. Cambridge, Mass.: Harvard University Press.

Cartwright, Timothy A., & Brent, Michael R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. Cognition, 63(2), 121-170.

Chomsky, Noam. (1965). Aspects of the theory of syntax. Cambridge, Mass.: M. I. T. Press.

Gomez, Rebecca, & Maye, Jessica. (2005). The Developmental Trajectory of Nonadjacent Dependency Learning (Vol. 7, pp. 183 - 206): Psychology Press.

Harris, Z. S. (1951). Structural linguistics. Chicago, IL: University of Chicago Press.

MacWhinney, Brian. (2000). The CHILDES Project: Tools for Analyzing Talk. Mahwah, NJ: Lawrence Erlbaum Associates.

Maratsos, Michael P., & Chalkley, Mary A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In Nelson, K. E. (Ed.), Children's Language (Vol. 2). New York, NY.: Gardner Press.

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

Mintz, Toben H. (2006). Finding the verbs: distributional cues to categories available to young learners. In K. Hirsh-Pasek, R. M. Golinkoff (Ed.), Action Meets Word: How Children Learn Verbs (pp. 31-63). New York: Oxford University Press.

Mintz, T. H. (2007). Category Induction from Lexical Co-occurrence Patterns in Artificial Languages. Invited presentation, Current Issues in Language Acquisition: Artificial & Statistical Language Learning, University of Calgary. Calgary, Alberta, June.

Mintz, Toben H., Newport, Elissa L., & Bever, Thomas G. (1995). Distributional regularities of grammatical categories in speech to infants. Paper presented at the 25th Annual Meeting of the North Eastern Linguistics Society, Amherst, Mass.

Mintz, Toben H., Newport, E. L. , & Bever, Thomas G. (2002). The distributional structure of grammatical categories in speech to young children. Cognitive Science, 26, 393-424.

Pinker, Steven. (1987). The bootstrapping problems in language acquisition. In MacWhinney, B. (Ed.), Mechanisms of language acquisition. New York, NY.: Springer-Verlag.

Redington, Martin, & Chater, Nick. (1997). Probabilistic and distributional approaches to language acquisition. Trends in Cognitive Sciences, 1(7), 273-281.

Redington, Martin, Chater, Nick, & Finch, Steven. (1998). Distributional information: A powerful cue for acquiring syntactic categories. Cognitive Science, 22(4), 425-469.

Sachs, Jacqueline. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. In Nelson, K. E. (Ed.), Children's language (Vol. 3). Hillsdale, NJ.: Lawrence Erlbaum Associates, Inc.

Saffran, Jenny R., Aslin, Richard N., & Newport, Elissa L. (1996). Statistical Learning by 8-Month-Old Infants. Science, 274(5294), 1926-1928.

Suppes, Patrick. (1974). The semantics of children's language. American Psychologist, 29 (2), 103-114.

Theakston, Anna L., Lieven, Elena V. M., Pine, Julian M., & Rowland, Caroline. (2001). The role of performance limitations in the acquisition of verb argument structure. Journal of Child Language, 28(1), pp. 127-152.

Toutanova, Kristina, Klein, Dan, Manning, Christopher, & Singer, Yoram. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. Paper presented at the HLT-NAACL 2003.

Toutanova, Kristina, & Manning, Christopher D. (2000). Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. Paper presented at the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, Hong Kong.

Colliding Cues in Word Segmentation: The Role of Cue Strength and Individual Differences

Daniel J. Weiss, Chip Gerfen, and Aaron Mitchel
The Pennsylvania State University

1. Introduction

Research over the past decade has yielded significant progress in our understanding of how domain-general learning mechanisms contribute to early language acquisition by tracking statistical dimensions of the input (e.g., Saffran, Aslin & Newport, 1996; Saffran, Newport & Aslin, 1996; Maye, Werker, & Gerken, 2002). Nevertheless, there remains a surprising lack of concurrent growth in research on how general cognitive abilities interact with language acquisition. This may derive from the fact that, individual variation in cognitive abilities notwithstanding, language acquisition itself is ubiquitous in humans. At the same time, it is also well-known that the course of acquisition does vary considerably across individuals (Bates, Bretherton, & Snyder, 1991), and we hypothesize that such variability may be crucially linked to general cognitive abilities. Few studies directly support theoretical accounts connecting general cognitive functioning with early stages of language acquisition e.g., Morgan & Saffran, 1995; Lalonde & Werker, 1995). Here, we test the issue by correlating performance on a colliding cue word segmentation task (see below) with performance on a Simon task, a non-linguistic cognitive task typically thought to involve executive processes such as inhibitory control and selective attention (Simon & Small, 1969; see Lu & Proctor, 1995 for an overview).

1.1 Statistical learning in Speech Segmentation

A central challenge of early language acquisition is segmenting words from the running stream. This challenge is amplified by the lack of invariant cross-linguistic cues to guide the learner to correctly segment the input stream (Klatt, 1979). Despite the continuous nature of speech, however, experienced language