

Cross-Cultural Differences in the Input to Early Word Learning

Jesse Snedeker (snedeker@wjh.harvard.edu)

Peggy Li (pegs@wjh.harvard.edu)

Sylvia Yuan (yuansy@wjh.harvard.edu)

Department of Psychology, 33 Kirkland St.

Cambridge, MA 02138 USA

Abstract

Young English learners typically have vocabularies dominated by nouns, while their Mandarin learning counterparts have relatively more verbs. Novice language learners must initially learn words from the situational contexts in which they appear, since they have limited access to syntactic or semantic context. These studies explore whether differences in these situational contexts could account for the differences in early vocabulary composition. We assessed the potency of this word-to-world mapping procedure in each language by asking adults to identify nouns and verbs from their extralinguistic contexts in maternal speech. The full pattern is examined in English native speakers given either English or Mandarin input (Exp. 1) and Mandarin native speakers given English or Mandarin input (Exp. 2). Results indicate that adults identify more nouns than verbs in English input, but identify as many verbs as nouns in Mandarin input. Thus cross-linguistic differences in the information that is available in extralinguistic contexts may account for some of the differences in early vocabulary composition.

Introduction

A child's vocabulary is the product of a constantly evolving interaction between the child and her experiences with language. One of the primary challenges for research on lexical development is unraveling the respective contributions of the learner and the input and tracing how this interaction changes over time. Crosslinguistic investigation of vocabulary composition is critical to this endeavor. By observing children whose language experiences vary, we can map out both how the input shapes the lexicon and the ways in which children are resistant to variation.

For twenty years, research in this area has centered on Gentner's noun dominance hypothesis (1982). Gentner compared parental reports of children's early vocabularies in a variety of languages. She found that nouns always made up the majority of the child's first words, even in languages where verbs are perceptually salient and arguments may be omitted. She argued that nouns are over-represented, relative to their frequency in input, because they refer to categories that are conceptually and perceptually simpler. This claim has been challenged by researchers who maintain that children learning Mandarin (Tardif, 1996), Korean (Choi & Gopnik, 1995), and Tzeltal (Brown, 1998) have early vocabularies in which the number of verbs equals or surpasses the number of nouns.

Our goal is not to judge whether language plays a role in shaping vocabulary composition. Nor is it to determine whether person and object labels are more easily learned regardless of language. These two positions are both well-evidenced and compatible: Studies which have examined the speech of children and adults have generally found both an effect of language and a difference between the composition of the input and the child's vocabulary (Tardif, Shatz & Naigles, 1997; Tardif, Gelman & Xu, 1999; for a review see Gentner & Boroditsky, 2001). Our goal instead is to explore one facet of the input that might favor or hinder words from different classes.

Previous research on crosslinguistic differences in vocabulary composition has focused on two ways in which the input can shape children's lexicons. First, many researchers have examined the frequency of different types of words in the input, making the plausible assumption that the number of learning opportunities should effect the likelihood that the child acquires a word. Mandarin, Korean and Tzeltal, the languages argued to show early verb dominance, all allow subject and object omission. Consequently, casual speech in these languages includes many utterances that contain a main verb but no nouns. Children learning a language of this kind are exposed to more verb tokens and fewer noun tokens than children learning English (Tardif et al., 1997; Choi, 2000).

The second type of explanation places the explanatory weight on the perceptual salience of nouns and verbs in connected speech. Presumably a child can only learn a word if she can segment the word-form from the ongoing speech stream. Previous work suggests that this task is easier when the word appears at the beginning or end of the utterance (Fernald, McRoberts & Herrera, 1992; Newport, Gleitman & Gleitman, 1977). Tardif and colleagues (1997) analyzed child-directed speech in Mandarin, Italian, and English to explore whether differences in the typical position of nouns and verbs in each language might account for the observed variation in vocabulary composition. They found that in Italian and English, languages where nouns dominated the early lexicon, verbs were generally buried away in the center of the utterance while nouns grabbed the salient utterance final position. In contrast, Mandarin speaking caregivers were more likely to produce sentences that ended with a verb (for parallel findings in Korean see Au, Dapretto & Song, 1994; Choi, 2000).

This work has been useful in understanding crosslinguistic variation in vocabulary composition. However, by focusing on the frequency and distribution of

word classes, researchers have overlooked another way in which language could shape word learning. Learning a word does not consist of merely of hearing the word and isolating it from speech. To learn a word a child must also pair the word-form with its meaning. Factors that influence the child's ability to perform this mapping, should also affect vocabulary composition and are therefore a potential source of crosslinguistic variation.

Early word learning is limited by the child's initial representation of the input. Because novice language learners know few words and little syntax, they must initially learn words from the real-world contexts in which they occur. Thus we would expect that children's early vocabularies would be limited to words whose meanings can be identified solely from the situational concomitants of their use. To the extent that languages vary in their stock of common nouns and verbs or in the pragmatic and situational cues that are available for word learning, early vocabularies should also vary.

The Human Simulations

Gleitman and colleagues have conducted a series of experiments which illustrate the potential importance of the mapping problem for explanations of vocabulary composition (Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker, Gleitman & Brent, 1999; Snedeker & Gleitman, in press). They find that developmental changes in vocabulary composition within a single language (English) can be explained by changes in the child's ability to use linguistic context to solve the mapping problem. In these studies adult subjects try to identify words from partial information about the contexts in which they occur in infant-directed speech. In some conditions the adults are given only the extralinguistic context, information that is available to novice language learners. In other cases they are provided with linguistic information, like syntactic context or co-occurrence, which would only be available to savvy learners.

Conceptually, these experiments are analogous to computer simulations in which a device, endowed with whatever ("innate") ideas and learning procedures its makers program into it, is exposed to data of the kind naturally received by the target learner it is simulating. The measure of success of the simulation is how faithfully it reproduces the learning function for that target using these authentic data. These experiments provide us with an estimate of the psychological potency of the cues to word meaning that are available in the real learning situation.

In prior work, we adopted this paradigm to determine whether the quality of the information in word-to-world pairs was sufficient to account for the composition of early vocabularies in different linguistic environments (Snedeker & Li, 2000). The subjects in these studies attempted to learn words by watching silent videotapes of the extralinguistic contexts in which a word occurred in either Mandarin or English input. In one experiment adult Mandarin speakers were shown videos of the 24 most

frequent nouns and verbs from a sample of Mandarin infant-directed speech. Their performance was compared with a parallel study in American English (Snedeker et al, 1999). We found a reliable interaction between language and word type. Verbs were identified more often in Mandarin while nouns were identified more often in English. There was also a reliable difference in the response biases of the Mandarin and American subjects, Mandarin subjects were more likely than American subjects to give verb responses regardless of the syntactic category of the target word.

This difference in bias illustrates a critical flaw in the design of the previous study: the input language and subject population were perfectly confounded. Mandarin adults viewed the tapes from Mandarin speaking mothers and American subjects viewed tapes from the English speaking mothers. Thus the observed differences in noun and verb identification could reflect either differences in the input that children receive in the two language environments or differences in the strategies and biases that the adult word learners bring to the task. Verb performance in the Mandarin study may have been better because the Mandarin mothers were more likely to use verbs with observable correlates, or performance may have been better simply because the Mandarin speaking adults had a bias to assume that all target words were verbs.

The current paper addresses this learner-input confound by fully crossing the subject population with the input language. In Experiment 1, we asked English-speaking students to identify words based on the situations in which they occur in Mandarin input. The results of this experiment were compared with those of a parallel study in which American students were given input from English speaking parents (Experiment 1 from Gillette et al, 1999; henceforth GGGL). In Experiment 2, we presented the same videotapes of English and Mandarin input to Taiwanese students.

These experiments have two goals. The first is methodological. Researchers using the human simulation paradigm have argued that it provides a glimpse of the information that is available in a particular representation of the input (see Gillette et al., 1999; Snedeker & Gleitman, in press). This argument is supported by examining the parallels between the simulations and the pattern of lexical development in children. In the current experiments we attempt to validate that argument by directly examining the sensitivity of this paradigm to manipulations of the input and the learner. If the paradigm primarily reflects the information that is provided in the videos then we would expect similar patterns of performance from the American and Taiwanese students. Our second goal is to explore whether differences in the contexts of word use could account for cross-linguistic variation in children's early vocabularies. English speaking children learn proportionally more nouns than their Mandarin speaking counterparts. If these differences are partially attributable to differences in the difficulty of the mapping problem, then we would expect that all our adult subjects would show a stronger effect of syntactic category when given the English input.

Experiment 1

In this experiment English-speaking adults watched silent videos of Mandarin-speaking mothers playing with their children. They were asked to guess what word the mother said each time a tone was sounded. Half of the target words were nouns and half were verbs. On every trial subjects were told whether the target word was a noun or verb. This was done to eliminate the differences in response bias that were found in Snedeker & Li (2000). The results of this experiment were compared with those of GGGL, where English-speaking adults were presented with videos of English-speaking mothers playing with their children.

Because the subject population is held constant in this comparison, we should be able to observe whether it is the input that contributes to differences in the kinds of words ‘acquired’ by adults. If input does play a role, then these adults should display different patterns of performance in noun and verb identification given different language input. In contrast, if it is not the input, but the biases or strategies of the speakers that drove the differences we observed in previous study, then we should expect to see the same pattern of performance by these English speakers in the two input conditions.

Methods

Participants 36 students at Harvard University participated. All were native speakers of English.

Stimuli The stimuli were drawn from videotapes of mothers playing with their 18-24 month old children. The six tapes, each approximately one-hour long, recorded 2 boys and 4 girls with their respective mothers playing with a set of toys provided by the experimenter. The mother was asked to “play naturally” with her child. The situations in which the children were taped, the toys that were provided and the instructions that were given to the mother were based on those used by GGGL.

The procedure for selecting the target words was identical to that used by GGGL. The videotapes were transcribed and the 24 most common nouns and verbs used by the mothers were chosen as targets. Table 1 shows a complete list of target nouns and verbs. We divided both the noun and verb targets into 4 frequency groups. Three presentation lists were constructed, each containing two words randomly selected from each frequency group.

For each word, six instances in which the word was used were randomly selected for inclusion in the study. For each instance, a video clip was constructed that began 30 seconds before the target word was used and ended 10 seconds after it was said. In many cases, the mother said the target word at another time during this 40-second period. In these cases, the clip was expanded to include 30 seconds before the first use and 10 seconds after the last. Each of the uses of the word in a single clip counted as one of the six stimuli. The audio was removed from the video clips and a tone was inserted exactly where the word had been.

Table 1: Target Words in Mandarin Input

Noun	English Meaning	Verb	English Meaning
qiu2	ball	chui1	blow
che1	car	jiao4	call
quan1quan1	circle	guan1	close
bing3gan1	cookie	lai2	come
bei1	cup	he1	drink
gou3	dog	che1	eat
walwa1	doll	fei1	fly
ya1ya1	duck	na2	get
xiang4	elephant	gei3	give
yan3jing1	eye	qu4	go
ba4ba	father	you3	have
shou3	hand	kan4	look/see
mao4	hat	kai1	open
tou2	head	wan2	play
ma1ma	mother	tui1	push
ming2zi4	name	fang4	put
jie3jie	older sister	jiang3	say
ren2	people	chang4	sing
fei1ji1	plane	zou4	sit
dian4hua4	telephone	shuo1	speak
dong1xi1	thing	zuan3	turn
wan3ju4	toy	zou3	walk
shui3	water	yao4	want
mei4mei	younger sister	hui4	will

Procedure Subjects were tested in groups of one to three. They were told that they were going to watch some clips of mothers playing with their children but were not told that the mothers in the videos were speaking Mandarin. They were asked to write down their best guess of the word the mother was saying each time they heard a tone. After hearing six tones, corresponding to six maternal utterances of the same word, they were asked to reconsider all the input and offer a final guess. The subjects were told whether the target was a noun or a verb. This procedure was repeated for 16 different words, half of them nouns and half verbs.

Coding A response was coded correct if it was the direct English translation of the target word. These judgments were made by a bilingual speaker. 94% of the words had a single English translation equivalent. The remaining cases were either ones in which a single English word is used in place of two Mandarin words (“sister” for both “jie3jie” and “mei4mei”) or a single Mandarin word has the meaning of two English words (“kan4” for “look” or “see”). In these cases a permissive coding criterion was adopted.

Results

Figure 1 shows the percentage of correct responses on final trial for English speakers in both input conditions. The results for English language input are taken from GGGL. As we noted earlier, when asked to identify words solely from situations in which they are used by American parents, American college students identify many of the nouns but do poorly on most of the verbs, resulting in a strong effect of Syntactic Category on final responses ($M = 45\%$, $M = 15\%$ for nouns and verbs respectively; $F=12.57$, $p < 0.001$).

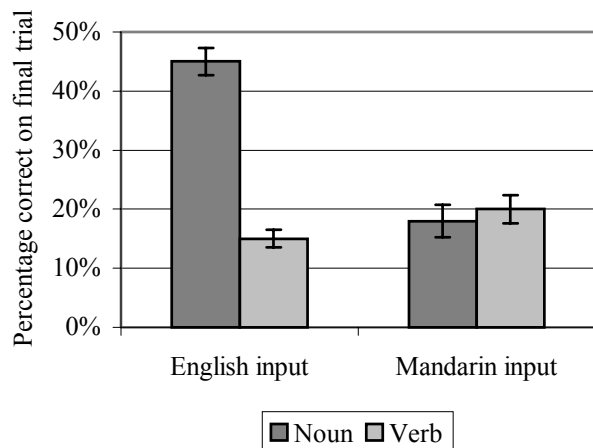


Figure 1: Correct Identification by English Speakers

The right side of Figure 1 illustrates the performance of English speakers who are asked to “identify” words from the situations in which they were used by Mandarin-speaking mothers. When given this input the American students find it equally easy to identify words from the two syntactic categories ($M = 18\%$ for nouns, and $M = 20\%$ for verbs; $F=0.483$, $p > 0.4$).

To compare the performance of the subjects in the two input conditions we performed an ANOVA with one within subject variable (Syntactic Category) and one between subject variable (Input Type). There was a strong interaction between Input and Syntactic Category ($F=40.310$, $p < 0.001$), confirming that the discrepancy between noun and verb identification varied with the language input. American students were better at identifying nouns but only when they were given English input. The dramatically high performance in noun identification for this group raised the overall performance for English input and overall noun identification across the groups, resulting in main effects of both Syntactic Category ($F=79.697$, $p < 0.001$) and Input ($F=22.164$, $p < 0.001$).

Experiment 2

A parallel experiment was conducted with Mandarin-speaking adults. In addition to showing Mandarin tapes to Mandarin speakers, we also showed English tapes to another group of Mandarin speakers. This gives us a second opportunity to examine the effect of changing input within a single population and ensure that the effects of input

language in the first experiment are actually attributable to differences in the input rather than differences in the subjects’ cultural knowledge or sense of identification. Furthermore, a comparison between the two experiments allows us to determine whether there is any interaction between input and the subject population being tested.

Methods

Participants 84 students from the National University in Taiwan participated in the experiment (42 in each input condition). All were native speakers of Mandarin.

Stimuli and Procedure We used the videotaped stimuli from Experiment 1 for the Mandarin Input condition and the videotaped stimuli from GGGL for English Input condition. Table 2 shows the complete list of target words for English tapes. The procedure was identical to Experiment 1 with the exception that the written instructions and debriefing were in Mandarin.

Table 2: Target words in English input (from GGGL)

Noun		Verb	
bag	music	come	play
ball	nose	do	pop
camera	peg	fall	push
daddy	people	get	put
drum	pig	go	say
elephant	pilot	hammer	see
hammer	plane	have	stand
hand	shoes	know	think
hat	swing	like	throw
hole	tail	look	turn
kiss	things	love	wait
mommy	toy	make	want

Coding For the Mandarin input condition, a response was coded correct if it contained the target word. For the English input condition, a response was coded correct if it contained the direct Mandarin translation of the target word. These judgments were made by one bilingual speaker. 92% of the English target words had a single translation equivalent. In the cases where there was no single translation equivalent, we applied the same permissive coding criterion that had been used in Experiment 1.

Results

The results for Mandarin speakers were quite similar to those for English speakers. Figure 2 shows the percentage correct on final trial for the two input conditions. Like English speakers, Mandarin speakers who were given Mandarin input performed equally well on noun and verb identification ($F=1.872$, $p > 0.15$), whereas those given English input identified more nouns than verbs ($F=53.367$, $p < 0.001$). As with the English speakers, a comparison

between the two Mandarin-speaking groups revealed a significant interaction between Input and Syntactic Category ($F=10.542$, $p < 0.005$), indicating that the discrepancy between noun and verb identification indeed varied with the input given. The high rate of identification for nouns in the English Input contributed to a main effect of Syntactic category ($F=23.807$, $p < 0.001$).

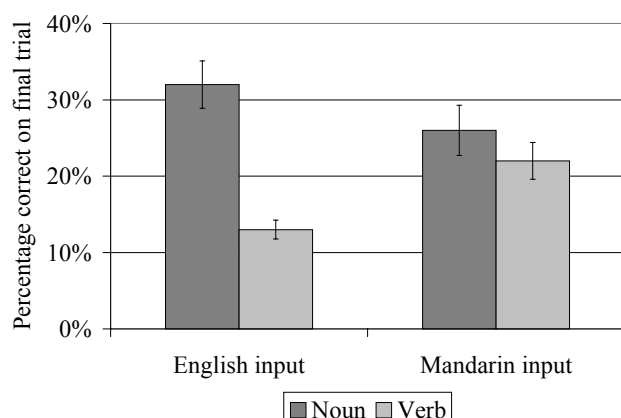


Figure 2: Correct Identification by Mandarin Speakers

Thus the results of both experiments indicate that when we hold the subject population constant, there are reliable effects of input language on the relative proportions of nouns and verbs that are identified. Namely, when given English input subjects correctly identify more nouns than verbs whereas when given Mandarin input they perform equally well in identifying both types of words. Additional analyses were conducted to determine whether the subject population would have any reliable effect on the performance pattern. In other words, are the performance patterns solely input-driven, or do adult subjects also have an impact, depending on what language they have been speaking?

Comparing subject populations A comparison of English speakers and Mandarin speakers who received English input revealed a significant main effect of Speaker Population ($F=16.610$, $p < 0.001$), with English speakers identifying more words overall than Mandarin speakers. Similarly, an analysis of subjects who received Mandarin input also yielded a reliable main effect of Speaker Population, with Mandarin speakers doing better than English speakers ($F=4.831$, $p < 0.05$). In other words, given the same tapes, subjects whose native language matched the mothers' did better. This result could be explained in several ways. For example, subjects who were watching tapes of mothers speaking their native language might have been more comfortable guessing what the mothers were saying. Critically, the same performance pattern for noun and verb identification held for different groups of speakers, given the same language input. As the previous within-group analyses indicated, both English speakers and Mandarin speakers, upon receiving English input, identified more

nouns than verbs, but when given Mandarin input, performed equally well on identifying both types of words.

General Discussion

Our comparison of word-to-world mapping in Mandarin and English unearthed a pattern of performance across syntactic class that is similar to what we see in early vocabularies of young children. When confronted with the input to young children learning English, both American and Taiwanese college students were able to identify more nouns than verbs. In this environment, novice language learners appear to get powerful situational information about the meanings of nouns, but weak and misleading information about the meanings of verbs (Gillette et al., 1999; Snedeker & Gleitman, in press). In contrast, when they attempted to identify words from Mandarin infant-directed speech, they performed as well on verbs as they did on nouns. This result confirms the findings of our previous study and demonstrates that there is a difference in the information available to infants in the two environments.

We have conducted initial analyses to explore three possible explanations for these results. First, we examined the social-pragmatic cues that were available in the videotaped scenes. It has been suggested that American mothers spend more time labeling objects for their children. These labeling episodes are typically characterized by a direct gaze at the object being named and often include pointing. Across the two languages, we found that the direction of the child's gaze was a reliable predictor of performance for noun targets. The direction of the mother's gaze was not and the effect of pointing was marginal. Critically, we found no cross-linguistic differences in the frequency of these cues.

Second, we examined the semantic characteristics of the target words. In our previous work we found that the strong performance on English nouns, is largely due to subjects' ability to identify targets that pick out basic-level categories of objects or animals (e.g., *dog* or *cup*). Performance on superordinates (*thing*), relational nouns (*uncle*), and abstract nouns (*name*) is generally no better than performance on verbs (Snedeker et al., 1999). The Mandarin and English noun targets were semantically similar. In both cases basic-level object categories accounted for roughly half of the nouns. Verbs were categorized as observable (e.g., actions like *walk*) and unobservable (e.g., mental states like *love*). In both languages subjects performed considerably better on observable verbs (presumably because they are more likely to have reliable perceptual correlates). The Mandarin input contained greater proportion of observable verbs, accounting in part for relative improvement in performance for verbs as compared to nouns. But the differences between the two sets of target words cannot fully explain the cross-linguistic differences in noun and verb learning. If we limit our analysis to the targets that appear in both input conditions, the interaction between syntactic category and input language is still strong and reliable for both groups of subjects ($F's > 8.36$, $p's < 0.01$).

Finally, we looked at the referents of the target nouns to determine whether they were actually present in the scene, whether they could be clearly identified and whether they were acceptable members of that category. These measures correlated with noun performance. Mandarin mothers were far more likely than English mothers to use a noun to refer to an object that did not actually belong to the semantic category in question ($p < .05$). Often the mother used a word that is common in infant directed speech (e.g., “cookie”) in place of a word that is less frequent (e.g., “potato chip”). In the Mandarin input condition one-third of the targets noun sets included instances of lexical replacement. The adult subjects who saw these scenes, sometimes responded with the correct basic-level label for the referent or a superordinate term. If these responses are coded as correct, performance on noun targets improves in both Mandarin input conditions ($M=24\%$, $M=28\%$ for English and Mandarin speakers respectively) resulting in a marginal advantage for nouns over verbs (p 's $< .1$). In future research we will attempt to determine whether this difference in lexical replacements is a reliable feature of the two linguistic environments or an artifact created by the set of toys that were brought by the experimenter.

In this experiment we told our subjects the syntactic category of the target word, both because it results in a higher level of performance (making comparisons between conditions more sensitive) and because it would eliminate the differences in response bias between the Mandarin and English speaking adults (Snedeker & Li, 2000). But real infants are not given this information. Knowing the category that a word belongs to simplifies the word-to-world mapping problem. Furthermore, it does so asymmetrically: in English at least, syntactic category labels benefit nouns more than verbs (Snedeker et al., 1999). However, there is no reason to believe that our experimental findings depend on the provision of syntactic information. In Snedeker & Li (2000) category labels were not provided, yet we observed a similar pattern of performance (see introduction).

There are, undoubtedly, many reasons why children's vocabularies vary across languages and cultures. Prior analyses of the input have focused on frequency and perceptual salience. Here we explored the role that the mapping problem may play. Even if children hear a word often, segment it from the speech stream and establish a phonological representation, they still must identify its meaning by observing the contexts in which it is used. These experiments demonstrate that the difficulty of mapping a form to a meaning varies across word classes and across linguistic and cultural environments. The information available in extralinguistic contexts can account both for the predominantly nominal initial vocabulary of English learners and the more balanced vocabulary of Mandarin learners.

Acknowledgments

The ideas in this paper owe much to our conversations with Lila Gleitman and members of the Cheese Seminar (1996-

2000). Any loose ends or inaccuracies are our own doing. We are extremely grateful to Jane Gillette and Lila for sharing their data with us and allowing us to use their videotapes of English-speaking parents. We also thank Becky Huang for her assistance with coding and Carol Chan and Renee Lam for their help in preparing the videotapes of Mandarin-speaking parents.

References

- Au, Terry; Dapretto, Mirella & Song, You-Kyung (1994). Input versus constraints: Early word acquisition in Korean and English. *JML*, 33, 567-582.
- Brown, P. (1998). Children's first verbs in Tzeltal: Evidence for an early verb category. *Linguistics*, 36, 713-753.
- Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, 27, 69-96.
- Choi, S. & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, 22, 497-529.
- Fernald, A., McRoberts, G. & Herrera, C. (1992). Prosodic features and early word recognition. Paper presented at the 8th International Conference on Infant Studies, Miami, FL.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S.A. Kuczaj, II (Ed.), *Language development: Vol 2. language, thought, and culture*. Hillsdale, NJ: Erlbaum.
- Gillette, J., Gleitman, H., Gleitman, L., Lederer, A. (1999). Human simulation of vocabulary learning. *Cognition*, 73, 135-151.
- Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In C. Snow & C. Ferguson (Eds) *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press
- Snedeker, J., Gleitman, L. & Brent, M. (1999). The successes and failures of word-to-world mapping. In A. Greenhill, et al. (Eds.), *Proceedings of the 23rd Boston University Conference on Language Development*.
- Snedeker, J. & Gleitman, L. (in press). Why it is hard to label our concepts. To appear in Hall & Waxman (eds.), *Weaving a Lexicon*. Cambridge, MA: MIT Press.
- Snedeker, J. & Li, P. (2000). Can the situations in which words occur account for cross-linguistic variation in vocabulary composition? In J. Tai and Y. Chang (eds.), *Proceedings 7th International Symposium on Chinese Languages and Linguistics*.
- Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, 32, 492-504.
- Tardif, T., Gelman, S. & Xu, F. (1999). Putting the 'noun bias' in context: A comparison of English and Mandarin. *Child Development*, 70, 620-635.
- Tardif, T., Shatz, M. & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs : A comparison of English, Italian, and Mandarin. *JCL*, 24, 535-565.