Counting nouns and verbs in the input: differential frequencies, different kinds of learning?*

CATHERINE M. SANDHOFER, LINDA B. SMITH

Department of Psychology and Program in Cognitive Science,
Indiana University

AND

JUN LUO

Department of Computer Science and Program in Cognitive Science,
Indiana University

ABSTRACT

Previous research has focused on evaluating the nouns and verbs in parents' input through type/token ratios. This research offers an additional means of evaluating parent speech by first examining the frequencies of individual nouns, verbs and descriptors and second examining the learning task presented to children. Study 1 examines 25 transcripts from the CHILDES database of English-speaking parents' speech to children at five developmental levels ranging from 0;11 to 2;11 in age. Study 2 examines 50 transcripts from the CHILDES database of Mandarin-speaking caregivers' speech to children ranging from 1;9 to 2;3 in age. The results suggest that the patterns of frequency for individual nouns and individual verbs are different, but that the frequency patterns for nouns and the frequency patterns for verbs are similar in English and Mandarin. Further, this research suggests that in both languages the nouns in parents' input are similarly organized: the most frequent nouns spoken to children tend to name solid objects that share a similar shape. In contrast verbs' meanings in both languages tend to include more variable conceptual relations.

INTRODUCTION

The quantity and quality of language input influences language learning. For example, Lieven (1978) observed interactions between children and their parents and found that the individual differences in children's language productions reflect individual differences in parental input. Similarly, Goldfield (1993) found a significant correlation between noun types used by mothers and the noun types in their children's first 50 words. Also Huttenlocher, Haight, Bryk, Seltzer & Lyons (1991) reported a strong positive correlation between the number of words parents direct towards children and the number of words children later have in their lexicon. Moreover, they found that the specific words that were frequent in a parent's speech to a child were the same words acquired earliest by that child. All of these results tell us what seems obvious: a thorough understanding of language input will be crucial to explaining language learning in general and differences in language learning across individuals and languages. However the question of just how to measure language input is complicated (Richards, 1994) and controversial (Tardif, 1996; Gentner & Boroditsky, in press). In this paper we present evidence on one kind of complication: how the relevant measure of the input may depend both on what is to be learned and also on the underlying learning mechanism. We will be specifically concerned with nouns and verbs in parent's speech to children, a domain of current controversy. We centre particularly on the adequacy and meaning of type and token counts in evaluating language input to children.

The differential noun and verb input children receive from their parents has generated interest because of children's (at least English-speaking children's) differential acquisition of nouns and verbs. The early vocabularies of children contain proportionally many more nouns than verbs. Nelson (1973), for example, detailed some of the early language production of toddlers. Of the earliest words produced by children, she found 65% were object words, 13% were action words and only 9% were modifiers. Stern (1924) found that 78% of children's early word productions were common nouns, 22% were verb-like and none of children's productions were adjectives. Moreover, the MacArthur Inventory, a vocabulary checklist of the 680 most frequent words produced by young children (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994), contains 53% nouns, 15% action words and 9% descriptive words. Such consistent findings of a noun advantage in early vocabularies can be explained in two ways: 1) it reflects the input, that is, nouns dominate because children hear more nouns than other kinds of words or 2) it reflects biases in the child, that is a propensity for learning names for things before more relational terms.

Gentner (1982) argues against input frequency being the primary controlling factor in the creation of a noun advantage. She notes that in speech,

verbs present a relatively small number of types with a high token frequency. In contrast, she argues nouns present a large number of types with a low token frequency. She argues that according to simple learning theory, verbs should be learned before nouns because the number of individual tokens of individual words is higher for verbs than for nouns. Thus to explain the noun advantage, Gentner proposed the Natural Partition Hypothesis in which nouns dominate despite their disadvantage in the input because object categories are conceptually and perceptually simpler than the relational concepts of verbs, prepositions, and adjectives.

Several recent cross-language results have been presented as contradictory to Gentner's position. First children learning some languages (see Choi & Gopnik, 1995 on the acquisition of nouns and verbs in Korean; Tardif, 1996 on the acquisition of nouns and verbs in Mandarin) apparently do not show a noun bias in their early vocabulary development. Second, the languages in which a noun bias is not present or not as strong in children's early vocabularies appear to be languages in which nouns are relatively infrequent in the input. For example, Tardif, Shatz & Naigles (1997) examined caregiver speech to children in three different languages, English, Italian, and Mandarin. They measured the frequency of noun and verb types and tokens in both adult–to–child and child speech. They found that adult Mandarin speakers differed from English and Italian speakers in that the Mandarin speakers used less noun types and more verb tokens than either English or Italian speakers. This difference corresponded with different patterns of noun/verb frequency in children's production of Mandarin. The results are consistent with the idea that the relative frequency of nouns in the input determines their relative frequency in children's speech and their relative ease of acquisition.

All of these previous papers compare the input relevant to noun learning and the input relative to verb learning by the same metric. However, is the optimal input for noun and verb learning the same? One way to think about this problem is to think in terms of an unbiased general learner (e.g. connectionist networks) and what it might learn if given different patterns of inputs – many tokens of few types, the verb-like pattern, or many types with few tokens, the noun-like pattern. The likely outcome is that: (1) Many tokens of few types will lead to rapid learning of those few specific types but little generalization across types, and (2) Few tokens of many types will yield generalizations across types and thus rapid learning of these types IF THEY PRESENT A COMMON STRUCTURE and poor learning if each type presents a unique structure. Thus, which pattern of input is better seems likely to depend on precisely what must be learned and the structural similarities across types (see Gasser & Smith, 1998).

From this perspective, the optimal input for learning nouns and verbs may be very different. The common nouns that children encounter early may, as

names for concrete things, present a common semantic, conceptual, and syntactic structure. In particular, most early nouns may be count nouns that name concrete things in shape-based categories (see Rosch, 1975; Macnamara, 1982; Samuelson & Smith, 1999; Smith, Jones, Landau, Samuelson & Gershkoff-Stowe, 1999; Smiley & Huttenlocher, 1995). If this is so, then there is a sense in which any one noun is equivalent to any other noun in its contribution to children's general knowledge about how nouns map to object categories. Thus, for example, learning something about the range of items that are called *box* might reinforce or potentiate learning about the word *spoon*. In contrast, the verbs young children encounter early do not present equivalent semantic and syntactic structures. For example, the two common verbs *put* and *look* describe highly dissimilar actions and relations between arguments. Understanding the meaning of *put* would seem to provide little insight into the meaning of *look*.

In this study, we examine the type/token relations of nouns, verbs, and for comparison, descriptors in parent speech to children in English and Mandarin. We ask how instances of nouns, verbs, and descriptors are distributed in the input – many examples of a few particular words, what we will call a STEEP DISTRIBUTION or many examples distributed across many different words, what we will call a FLAT DISTRIBUTION. And we ask whether the distributional differences between the nouns and verbs in parents' speech are similar in the two languages. We also ask whether the many different nouns that children encounter are semantically similar. Correspondingly, we ask whether the many different verbs children encounter are also semantically similar. We address these questions by examining transcripts of parent speech to children from the CHILDES database.

STUDY 1: THE ENGLISH DATASET

We examined only transcripts that consisted primarily of mother–child conversations for the English dataset even though these transcripts were conducted by many different experimenters, for many different purposes, over many different age ranges. Our reasoning is this: this broad sampling of parent speech across many children should reflect the statistical regularities characteristic of the many diverse conversational contexts encountered by any one child. Thus in most of the analyses, we combine all the transcripts and analyse the resulting whole for what it can tell us about the structure of the word-learning environment encountered by children in general.

METHOD

Subjects

Dataset. Twenty-five mother–child transcripts were obtained from the CHILDES database (MacWhinney & Snow, 1990) for children from 0;11 to

2;11. Transcripts were selected to consist primarily of mother–child (as opposed to experimenter–child) conversations. The transcripts analysed are listed in the Appendix A.

Procedure

The words spoken by the caregivers and their frequencies were compiled into a list using the CHILDES program *freq*. Each word was then assigned to one of four categories: noun, verb, descriptor, or non-classified. The noun category included both concrete (e.g. *car*) and abstract nouns (e.g. *imagination*), but did not include pronouns. Similarly, the verb category contained both active (e.g. *jump*) and experiential (e.g. *think*) verbs but did not include auxiliaries such as *have* or *be*. The descriptor category contained both descriptive words and modifiers. Words that normally occupy multiple grammatical classes in which grammatical class can be assigned only on the basis of context, for example *dress*, were assigned by referring back to the transcripts to determine the context in which the word appeared. All other words, including, adverbs, pronouns and prepositions were not classified. A subset of the transcripts (322 types and 1432 tokens) were assigned to categories using the transcripts and context to determine the grammatical class. The results of these category assignments were compared to the category assignments using the *freq* method described above. Reliability was 89% agreement between the two assignment methods for types, and 92% agreement between the two assignment methods for tokens.

The resulting output was then handchecked and individual instances of regular noun and verb forms were combined. Irregular forms were not combined. That is, all instances of *cup* and all instances of *cups* were combined as tokens of a single type, but instances of *child* and *children* were counted as separate types. Similarly, within the verb category all regular usages of verbs were combined. For example, *throw* and *throwing* were counted as tokens of the same type, but the irregular form *threw* was counted as a separate type.

Developmental differences

To assess possible differences in the patterns of speech as a function of the developmental level of the child, we partitioned the transcripts into five categories as a function of age. The first age group contained five transcripts of children 1;3 or younger (range 0;11 to 1;3). The second age group consisted of five transcripts with children between the ages of 1;4 and 1;8. The third age group consisted of five transcripts with children between the ages of 1;9 and 2;1. The fourth age group consisted of five transcripts of children between the ages of 2;2 and 2;6. Finally, the last age group consisted of children between the ages of 2;7 and 2;11.

Adult judgments of nouns

To assess the common structure of nouns, the first two authors, native speakers of English, judged the most frequent nouns produced by parents using the semantic analysis described in Samuelson & Smith (1999). Forty-three of the 50 most frequent nouns were analysed; the excluded terms were proper names, character names or people. Each of the nouns was judged for syntactic category (count versus mass), the solidity of the referents (solid or nonsolid things), and the within-group similarity of their referents (in colour shape, or material).

To make judgments of within group similarity we asked a series of yes/no questions of the form: 'Are instances of this category generally similar in __?' The properties queried were shape, colour, and material. From these judgments each noun was designated as shape-based, colour-based, or material-based. Note that individual lexical items could be deemed, by this criterion, to be organized by more than one property or by no property. For example, by our judgment, the category *crayon* is both shape-based and material-based, whereas *cup* is shape-based, and *wind* is not organized by any of these properties.

Further, we judged whether the instances of each category were typically solid, nonsolid, or neither clearly solid or non-solid. We defined solid things as objects that have rigid shapes that they maintain when moved or that they return to when pressed or touched. We defined non-solid things as things that do not have rigid shapes that they maintain and do not return to their original shape when pressed or moved.

Finally, we judged whether each noun was a count or mass noun by thinking of how nouns are used in everyday conversation (not how they might be used). Our reasoning in judging the common usage as a count or mass noun is motivated by the complexity and fluidity of this syntactic distinction (see Bloom, 1996; Samuelson & Smith, 1999). We classified nouns using the following definitions: COUNT NOUNS can be preceded by the word *a* and numerals, have a plural form, and cannot be preceded by the word *much*. MASS NOUNS can not be preceded by the word *a* and numerals, do not have a plural form, and can be preceded by the word *much*. OTHER NOUNS included items like *cake* that can occur in both count and mass form or items like *glasses* that do not unambiguously fit the definition for count or mass noun.

Judgments of verbs

To assess the common structure of verbs, the first two authors judged the most frequent verbs produced by parents using the semantic analysis described by Naigles & Hoff-Ginsberg (1995). The 50 most frequent verbs

in the dataset were analysed. Each of the verbs was judged for kind of motion (internal state versus physical motion) and its actor–patient relationships (actor only verb, actor AND patient verb, or actor OR patient verb).

We made judgments about the kind of motion described by a particular verb by asking whether the activity described by that word involved physical motion that could be observed by another person. Words that described a physical activity were designated as PHYSICAL MOTION verbs and words that did not describe a physical activity were designated as INTERNAL STATE verbs. For example, the verb *go* describes a physical activity whereas the verb *want* describes an internal state.

We judged the actor–patient relationships by thinking of how the verbs are used everyday (and discounted specialized meanings such as 'running a subject.') We defined the actor/patient relations as the following (see Naigles & Hoff-Ginsberg, 1995 for further detail): ACTOR ONLY VERBS only the actor does the action, e.g. *go*. ACTOR AND PATIENT both the actor and patient are involved in the action, e.g. *give*. ACTOR OR PATIENT VERBS can involve both the actor and the patient, but one or the other of these roles can be omitted, e.g. *open*.

RESULTS

Types and tokens

Table 1 shows the number of words types and tokens for each of the three classified categories (noun, verb, and descriptor) for all of the transcripts combined. Overall, parents produced many more noun types than any other type, and over twice as many noun types as verb types. However, the number of tokens produced by parents shows similar frequencies for nouns and verbs. That is, individual verbs are repeated more often that individual nouns so although there are more noun types, there are just as many verb tokens as noun tokens. This discrepancy between noun and verb types and tokens in English has been previously reported in studies of both adult–adult and adult–child speech (Gentner, 1982; Johanssen & Hofland, 1989; Goldfield, 1993; Tardif, 1996).

Developmental trends in parent's speech?

We first examined the English-speaking caregivers' data to determine whether the patterns of noun, verb, and descriptor distributions change as a function of the experience of the language learner. To do so we found the relative frequency of nouns, verbs, and descriptors in parent speech by taking, for example, the number of nouns expressed as a proportion of the total nouns, verbs, and descriptors in parent speech. Figure 1 presents the

TABLE 1. *Number of types and tokens in the maternal speech of the English dataset*

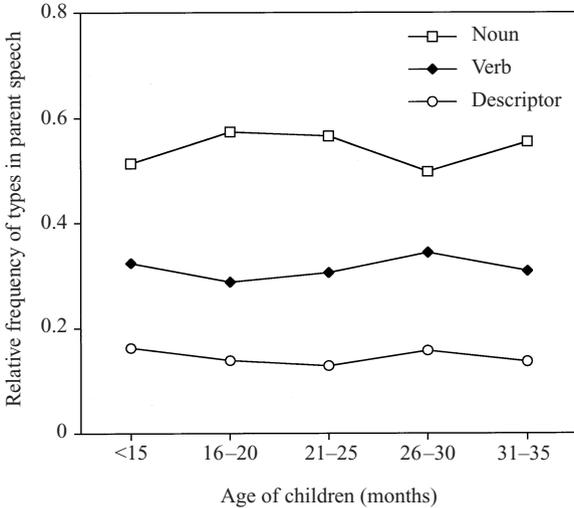|	Nouns	Verbs	Descriptors
Types	730	305	165
Tokens	4747	4595	938

Fig. 1. The relative frequency of noun, verb, and adjective types of English-speaking caregivers.

proportion of noun, verb, and descriptor TYPES in mother's speech to children at five different age levels. As can be seen, the relative frequency of noun types is somewhat higher than the relative frequency of verb types at all ages tested and this appears relatively consistent across the age ranges examined.

The relative frequency of noun, verb, and adjective TOKENS across the five age groups is shown in Figure 2. Here we see a somewhat different pattern. Again parents use substantially fewer descriptor tokens than noun or verb tokens, but in the first 4 of 5 age groups the relative frequency of noun tokens is slightly higher than the relative frequency of verb tokens. Mothers of children in the oldest age group, 2;7 to 2;11, do not share this pattern, but use more verb tokens than noun tokens. The number of verb tokens relative to noun tokens in parents' speech appears to increase with the age of the child.
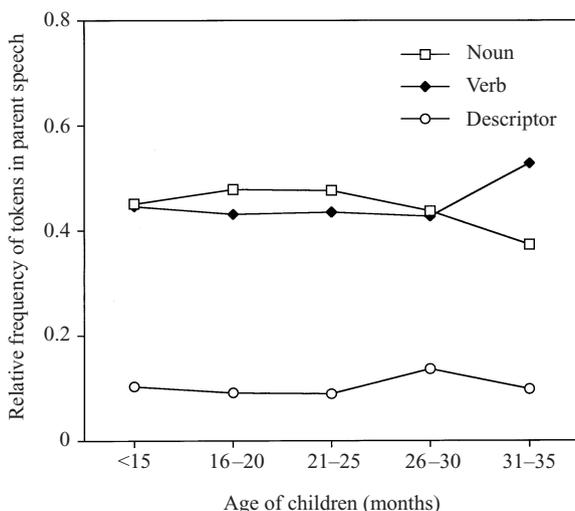
Fig. 2. The relative frequency of noun, verb, and adjective tokens of English-speaking caregivers.

High and low frequency words

Although the previous type/token frequencies provide an overall picture of the noun, verb, and descriptor input children experience in everyday speech with their parents, they provide neither a detailed picture of the statistical structure of these categories nor specific information on the particular nouns and verbs that children encounter. That is, although type/token ratios can point to the average frequency of nouns versus verbs, they offer little insight into how children learn about *spoon* versus *put*. Thus the remaining analyses will focus on detailing the most frequent nouns and verbs children encounter and analysing the statistical structure of the noun and verb categories.

In order to examine the statistical structure of instances of noun and verb distributions, we first partitioned noun and verbs into percentile categories. We did so by rank ordering each noun and verb in the combined corpus of noun and verb types by their token frequency. The ranked types were then partitioned by their percentile score. We examined four equally sized intervals of percentile categories: 33–49, 50–66, 67–83, and 84–100.[1] Within each percentile category the number of noun types for that category was divided by the total number of noun types for the corpus, and the number of verb types for that category was divided by the total number of verb types for

---

[1] The nouns and verbs in the 0–32 percentile category could not be further subdivided into smaller categories since the majority of types in this category occurred with an equal token frequency of 1.

the corpus. Thus, this measure provides information on how noun and verb types are distributed while controlling for the large difference in overall types of nouns and verbs

If parents produced an equal number of nouns with high, medium, and low frequencies and an equal number of verbs with high, medium, and low frequencies then this measure should not yield in differences between noun frequency at any percentile level or verb frequency at any percentile level. However, as shown in Figure 3 this is not the case. Although parents produce roughly equal numbers of nouns and roughly equal numbers of verbs in the 33–49, 50–66, and 67–83 percentile levels, at the highest percentile level, 84–100 parents are producing more verbs. That is, a disproportionate number of the verbs parents produce are used at very frequent levels.
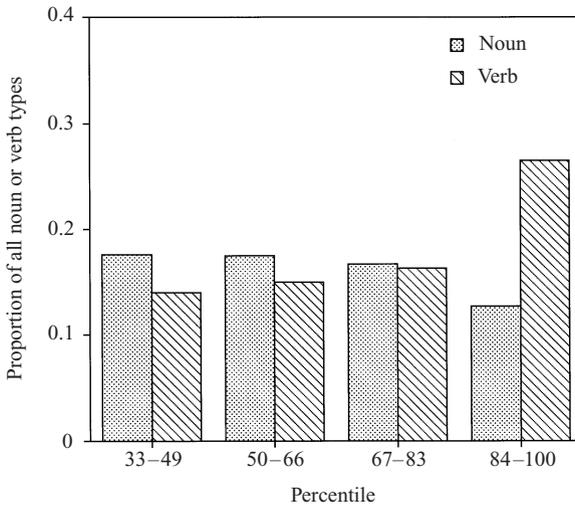

Fig. 3. Number of nouns as a proportion of all nouns and number of verbs as a proportion of all verbs in four percentile levels for English-speaking caregivers.

This pattern can also be seen by closer examination of the most frequently produced nouns, verbs and descriptors. These 50 most frequent nouns, verbs, and descriptors are listed in Appendix B. With the exception of the child's own name, parent's produce many nouns all with approximately equal low frequency. As seen in Figure 4 the token frequency of the 5th most frequent noun (kitty), the 25th most frequent noun (box), the 50th most frequent noun (doll) and the 100th most frequent noun (spoon) do not differ greatly. However, the frequency of verbs does not show the same pattern. There is a small set of verbs produced very frequently. Beyond this initial set
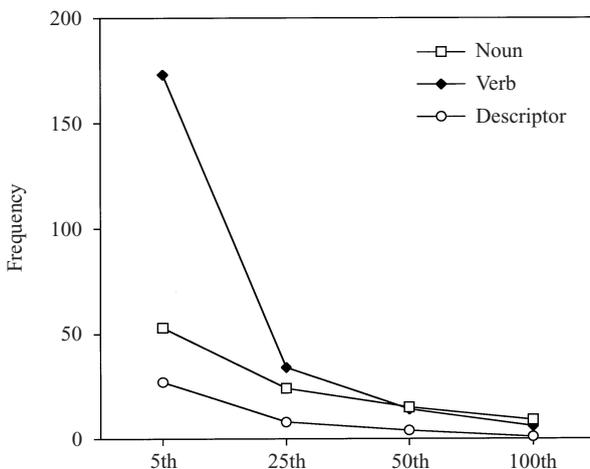
Fig. 4. The token frequency of the 5th, 25th, 50th, and 100th most frequent nouns, verbs, and adjectives produced by English-speaking caregivers.

of 10–15 verbs the token frequency of verbs declines such that the difference between the 5th most frequent verb (look), the 25th most frequent verb (hold), the 50th most frequent verb (jump, and the 100th most frequent verb (swim) is considerable. Finally, in contrast to nouns and verbs, only a few descriptors are used frequently in speech to children. Thus, the nouns in parent input present a flat distributions, lots of nouns, all of modest frequency, whereas the verbs present a steep distribution, a few verbs presented with high frequency and others presented with low frequency.

Structural commonalities among frequently produced noun types

Do the common nouns children encounter early present a common structure? Specifically, do the nouns children encounter frequently in speech represent concrete things of similar shape? We judged 74 % of the nouns to refer to solid objects, 58 % of these nouns to refer to objects in categories organized by shape, and 84 % of these nouns to be count nouns. As a whole, the most common nouns in parents' speech to children tend to be similarly organized. Indeed 72 % of these nouns are count nouns naming solid objects by shape. In sum, nouns that dominate parent speech name bounded (countable) objects by their perceptual properties.

Structural commonalties among frequently produced verb types?

We next asked whether the verbs children encounter also present a common structure. We judged 76 % of the 50 most frequent verbs to refer to physical motions. However, 20 % of these verbs were judged to be actor only verbs,

52 % were judged to be actor and patient verbs, and 14 % to be actor or patient verbs. Verbs thus seem more complicated and more diverse than nouns in meaning. Most nouns in the corpus refer to categories of perceptually similar things and in most cases to categories of things similar in shape. In contrast, verb meanings include more variable conceptual relations of actor and patient.

In total, the results of these analyses show that the type and token ratios for nouns and verbs match patterns previously reported in the literature for English (Johansson & Hofland, 1989; Goldfield, 1993; Tardif *et al.*, 1997; Gentner & Boroditsky, in press). Parents of young children use many more noun types than verb types, and the number of verb tokens used by parents is equal to or higher than the number of noun tokens used by parents. However, these analyses also provide information on the statistical structure of instances of nouns and verbs. Many noun types are produced with a relatively low frequency, a flat distribution, whereas the few verb types produced by parents are produced with high frequencies, a steep distribution. Moreover, we find that the most frequent nouns spoken by parent share a common semantic structure whereas the most frequent verbs spoken by parents are more diverse in meaning. The pattern fits that suggested earlier by Gentner (1982): early nouns are simpler and more uniform in their mappings to meanings and verbs are more relational, complex, and diverse. The statistical pattern of input that we observed seems to adaptively fit these differences – few tokens of lots of different (but semantically similar) nouns and many tokens of a few (semantically complex and individually unique) verbs.

STUDY 2: THE MANDARIN DATASET

The results of Study 1 suggest that the statistical distributions of nouns and verbs in English-speaking parents' speech to children differ, and that they do so in ways that may be related to the early noun advantage in English-speaking children. The goal of Study 2 is to examine a second language. Mandarin presents an interesting comparison case primarily because the evidence suggests that children do not show a marked noun advantage in early vocabulary production and because verbs – types and tokens – are more frequent in the input than are nouns (Tardif *et al.*, 1997). Moreover, Tardif *et al.* report that verbs appear in salient initial and final positions more often in Mandarin than in English (due to the pro-drop and null object features of Mandarin).

We ask whether, in addition to these differences, there are also differences in the distributional structure of noun and verb categories in the two languages. There are at least two ways that the Mandarin pattern could differ from the English pattern. First, the Mandarin and English patterns could

differ only in the relative numbers of verbs and nouns as previously reported, with verbs being more frequent than nouns. The distributional structures – steep for verbs, flat for nouns – may be the same. This would suggest that nouns and verbs share the same functions in the two languages. The second possibility is that the primacy of verbs in Mandarin could reflect more profound differences in the learning task presented to children. For example, if the frequency of verbs in the input and their primacy in language coincides with (or leads to) more transparent verb meanings, then the distribution of verbs in Mandarin might look more like that of nouns in English – lots of different verbs all used with modest degrees of frequency.

METHOD

Mandarin dataset

50 caregiver–child transcripts were obtained from the CHILDES database (MacWhinney & Snow, 1990) Beijing corpus (Tardif, 1993; 1996) for 10 children from 1;9 to 2;3. From these transcripts, we selected only the speech that was directed to the child. Although the English dataset contained only mother to child speech, these transcripts contain speech from multiple caregivers, as it is customary that many family members participate in child raising in Beijing.

Procedure

Since many words in Mandarin are homophones, the majority of the assignment to noun verb or descriptor category was accomplished by directly consulting the transcripts. However, the words spoken by the caregivers and their frequencies were also compiled into a list using the CHILDES program *freq* and this list was consulted for unambiguous cases. We assigned to one of four categories: noun, verb, descriptor, or non-classified as in the English dataset. We chose to classify instances of verb-adjectives as descriptors instead of as verbs for several reasons. Choi (1998) argues that since verb-adjectives comprise such a small proportion of the total verbs that their presence (or absence) should not affect the frequency counts of verbs as a whole. Second, the verb-adjectives in Mandarin are semantically similar to adjectives in English and we wished to provide a comparison for the English descriptor category. Lastly, since the assignment of these verb-adjectives is a somewhat arbitrary one, we chose the more conservative approach of counting them separately.

As in the English dataset, individual instances of regular noun and verb forms were combined. Irregular forms were not combined. For example, instances of *hai* and instances of *haizi* were combined as tokens of a single

type. Similarly, within the verb category all regular forms of verbs were combined. For example, *tiao* and *tiaole* were counted as tokens of the same type.

Judgments of nouns

The third author, a native speaker of Mandarin, judged 38 of the 50 most frequent nouns produced by Mandarin-speaking caregivers. The excluded terms were proper names, character names or people. Each of the nouns was judged for solidity (solid or nonsolid things), and within-group similarity (in colour, shape, or material) as in the English dataset. Note that Mandarin does not have a count/mass distinction. The judgments were made for objects as they appear in Beijing. That is, if all apples sold in Beijing were red in colour, the experimenter was to judge apples as being the same colour (even though he may have seen green apples in the U.S.)

Judgments of verbs

The third author also judged the 50 most frequent verbs produced by Mandarin-speaking caregivers by both the kind of motion and actor–patient relationships described in Study 1. The third experimenter remained blind to the results of the English judgments until after he had judged the Mandarin nouns and verbs.

RESULTS

Types and tokens

Table 2 shows the number of word types and tokens for each of the three categories (noun, verb, and descriptor). Overall, Mandarin-speaking care-givers produced more noun types than any other type. However, in striking contrast to English-speaking parents, Mandarin-speaking caregivers produce many more verb tokens than noun tokens – almost twice as many verb as noun tokens. Thus like in English, individual verbs are repeated more often that individual nouns, but the order of magnitude is much greater in Mandarin, with individual verbs being produced on average 44·4 times compared to 15·1 times in English. The dominance of verb tokens in Mandarin-speaking parent's speech has been previously reported. (Tardif, 1996, see also Choi 1998 for similar evidence from Korean).

High and low frequency words

We divided the Mandarin nouns and verbs into the four equal percentile categories: 33–49, 50–66, 67–83, and 84–100 using the same method as Study 1. Within each percentile category the number of noun types for that

TABLE 2. *Number of types and tokens in the caregiver speech of Mandarin dataset*

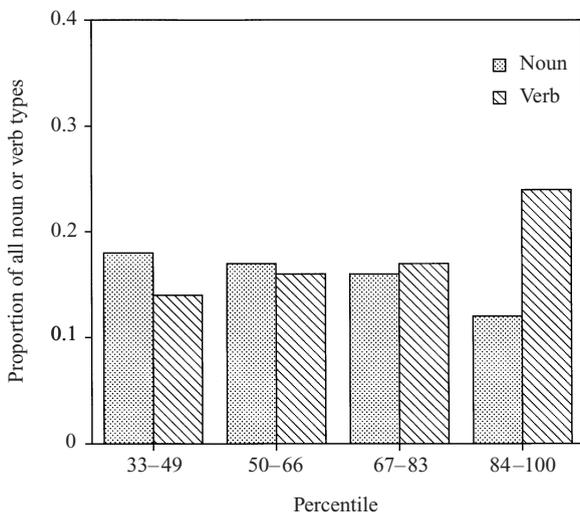|	Nouns	Verbs	Descriptors
Types	1041	664	301
Tokens	17097	29454	8218

Fig. 5. Number of nouns as a proportion of all nouns and number of verbs as a proportion of all verbs in four percentile levels for Mandarin-speaking caregivers.

category was divided by the total number of noun types for the Mandarin corpus, and the number of verb types for that category was divided by the total number of verb types for the Mandarin corpus.

Although the type/token ratios of nouns and verbs in Mandarin-speaking parents' input differ radically from the type/token ratios of nouns and verbs in English-speaking parents' input, the frequency distributions of nouns and verbs in English and Mandarin appear similar. As shown in Figure 5 the distribution Although parents produce roughly equal numbers of nouns and roughly equal numbers of verbs in the 33–49, 50–66, and 67–83 percentile levels, at the highest percentile level, 84–100 parents are producing more verbs. As in English, a disproportionate number of the verbs parents produce are used very frequently in speech to children.

Appendix C lists the 50 most frequent nouns, verbs, and descriptors in this corpus of Mandarin-speaking parents' speech in order of frequency. Again as
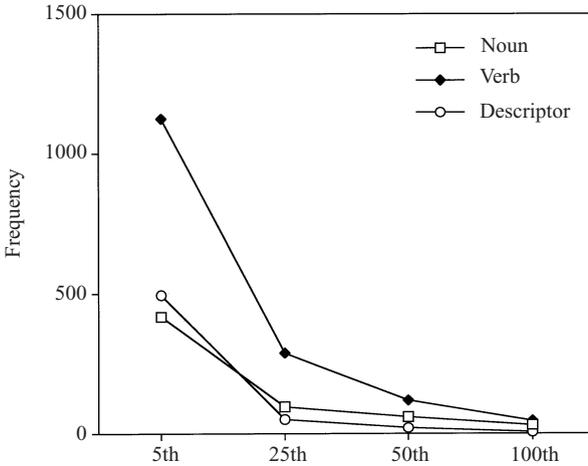
Fig. 6. The token frequency of the 5th, 25th, 50th, and 100th most frequent nouns, verbs, and adjectives produced by Mandarin-speaking caregivers.

in English, with the exception of the child's own name, Mandarin-speaking parents produce many nouns all with approximately equal modest frequency. As shown in Figure 6, the token frequency of the 5th most frequent noun (ba4ba), the 25th most frequent noun (yu2), the 50th most frequent noun (deng4zi) and the 100th most frequent noun (dian4shi4) do not differ greatly. And again as with English, there is a small set of verbs produced very frequently. Beyond this initial set of 10–15 verbs, the token frequency of verbs declines such that the difference between the 5th most frequent verb (deng3), the 25th most frequent verb (nong4), the 50th most frequent verb (diao4), and the 100th most frequent verb (jiu1) is considerable. Finally, in contrast to nouns and verbs, only a few descriptors are used frequently in speech to children. Thus similarly to English, the nouns in Mandarin-speaking parents' input present a flat distribution, lots of nouns, all of modest frequency, whereas the verbs present a steep distribution, a few verbs presented with high frequency and others presented with low frequency.

Structural commonalities among frequently produced noun types

We next asked whether the common nouns Mandarin-speaking children encounter early present a common structure? Specifically, do the nouns children encounter frequently in speech represent concrete things of similar shape? We judged 82 % of the nouns to refer to solid objects, 84 % of these nouns to refer to objects in categories organized by shape. As with English, the most common nouns in Mandarin-speaking parents' speech to children tend to be similarly organized – 79 % of these most frequent nouns name

solid objects by shape. Most nouns in the corpus refer to categories of perceptually similar things and in most cases to categories of things similar in shape.

Structural commonalties among frequently produced verb types?

We next asked whether the verbs children encounter also present a common structure. We judged 92 % of the 50 most frequent verbs to refer to physical motions. However, 14 % of these verbs were judged to be actor only verbs, 58 % were judged to be actor and patient verbs, and 28 % to be actor or patient verbs. Thus, although more of the most frequent Mandarin verbs refer to physical motions, the Mandarin verbs still seem more complicated and more diverse than nouns in meaning. That is, verb meanings in English and Mandarin include the more variable conceptual relations of actor and patient.

GENERAL DISCUSSION

The present results have implications for three issues: (1) differences be-tween nouns and verbs in the input and the relevance of these differences to how children acquire early nouns and verbs, (2) similarities and differences in the input presented by parent speech for children learning English and by parent speech for children learning Mandarin, and (3) the natural partitions hypothesis. We consider each of these in turn.

Nouns versus verbs

These two studies show that there are differences between noun and verb input that go beyond merely types, tokens, and type–token ratios. The distributional differences observed here raise the question of whether type–token counts are the best metric for assessing the information in the input relevant to acquisition. Clearly, all learning theories would predict that the frequencies of individual words in parent input should be reflected in children's productions. However, there is no straightforward prediction from learning theory as to whether a pattern consisting of many types (and few tokens of each) or a pattern consisting of many tokens (of a few types) is optimal. Instead it seems likely that the right metric for acquisition depends on just what is to be learned. Accordingly, in the present studies, we also attempted a small-scale analysis of how nouns and verbs map to meaning.

The results show that the input presents a flat distribution of nouns – many types all presented with modest frequency. In brief, individual nouns do not differ much in how often they occur in speech to children. They also do not differ much in the kinds of things to which they refer. Most of the nouns in parents' speech refer to solid objects that are similar in shape. It seems likely that these two facts – the flat distribution and the common category structure – are related. We propose that the distributional pattern is

577

flat because early object categories are all roughly equal: names for concrete things recognizable by their perceptual properties. If this is so, then the flat distributional pattern characteristic of nouns may be sufficient, indeed optimal for rapid learning. Children may easily learn many different object names precisely because the categories described by common nouns are similarly organized. Children may not need many repeated experiences with each individual noun to understand how nouns map to object categories. Instead, they may induce a general principle about likely noun meanings that they can apply even to rarely encountered object names.

These ideas about early nouns suggest that both children and their parents may genuinely treat object names as an open class – all objects are nameable things. This proposal fits the evidence in the literature from children learning English: at the time of the early vocabulary spurt children commonly ask for the names of things as if they expect all things to have distinct names (see Anglin, 1977; MacNamara, 1982; Markman, 1989 on this point). Moreover, when told the name of a novel single object, children appropriately generalize the name to new instances (Markman & Hutchinson, 1984; Golnikoff, Mervis & Hirsh-Pasek, 1994; Waxman, 1994; Smith, in press). All this suggests that even though individual noun types appear infrequently in the input, the CORPUS of nouns refers to similar enough entities that children can easily learn names for things.

In contrast, the present results show that the input presents a steep distribution of verbs – many tokens of a very few types. Parents use a small set of verbs repeatedly in many contexts and other verbs infrequently. In brief, individual verbs are not equally available for learning from parent speech. And, individual verbs, particularly the common early ones, differ richly and intricately in meaning and argument structure (Gleitman, 1993; Goldberg, 1998). We see this as well in our admittedly modest analysis of verbs in the present corpus. The verbs in parents' speech studied here differ in whether they refer to a physical motion that involves only an actor, an actor and patient, or either an actor OR patient. As Gleitman (1993) suggested, children may need repeated exposure to the same verb in multiple real world contexts and in multiple linguistic contexts that present different argument structures to infer the meaning of that particular verb (see also Goldberg, 1998). Again, we speculate that the steep distribution of verbs in the input and the intricacies of verb semantics may be related. And again, this idea is supported by evidence on how children learn verbs. We know of no cases in which anyone has reported a child asking for the verb that refers to some experienced or observed event. Indeed, we cannot even imagine what such a case would be like. Further, although young children do map novel verbs to novel events they do so conservatively apparently on a case-by-case basis (Tomasello, 1992). The steep distributional pattern of verb input may be optimal for learning in depth about a small set of verbs. Such narrow but

deep learning may be necessary to induce broader generalizations about the system of verb meanings. Initially dense presentations of a few verbs in multiple linguistic contexts may be optimal for learning patterns of argument structure and ultimately many verbs.

These suggestions that the distributional patterns of nouns and verbs may be optimal for learning their respective meanings is not meant to imply that parents strategically choose to use many nouns with equal frequency but a few verbs with higher frequency. Rather, we suspect that these patterns reflect deeper truths about the linguistic functions of nouns and verbs: nouns refer, pointing to things in the world or abstract ideas; verbs relate, integrating and conjoining meanings to create a proposition (Goldberg, 1998). In this context, it is interesting to note that the distributional patterns are highly similar in both English and Mandarin, despite the considerable differences in the linguistic structure of the two languages (Tardif, 1996). We conjecture that the distributional patterns of nouns and verbs that we observed here will also be found in adult–to–adult speech in both languages. This is expected if the patterns reflect deep truths about the linguistic functions of nouns and verbs and not merely strategies for talking to children.

Before turning to cross-language issues, we comment briefly on the findings concerning descriptors, or adjectives. These were not the main focus of this study and were counted primarily because they have sometimes been classified as verbs in Mandarin. Therefore we offer only preliminary comments about their distribution and meaning. On the surface the most coherent description of adjectives may be that they describe properties of objects. But the types of properties they describe are diverse: perceptible properties of objects e.g. *stinky*, *soft*, or *salty*, evaluative descriptions, e.g. *good*, descriptions of internal states, e.g. *sleepy*, descriptions of actions rather than objects, e.g. *slow*, and also descriptions of relations, e.g. *empty*. Furthermore, as is readily apparent, the relative frequency of adjectives in parent speech in both languages is considerably lower than nouns and verbs. This low frequency coupled with the diversity and lack of systematicity across adjective meanings may be related to the relative lateness of adjective acquisition (Carey, 1978, 1982; Gasser & Smith, 1998).

English versus Mandarin

We included a comparison of Mandarin and English because prior research shows differences in the relative frequency of noun and verb types in children's productions and in noun and verb tokens in parent speech. These past findings show a clear connection between input and children's speech: children learning English hear more nouns than verbs and say more nouns than verbs, whereas children learning Mandarin hear more verbs than do English-speaking children and they say more verbs. However, as Tardif

(1996) notes the correspondence between input and output in these studies is imperfect. The differences in noun and verb type/token ratios are greater in the parent input than in the children's productions. The task of language learning somehow tempers the differences in input. We believe the present results may help may help explain this fact. The effects of type and token frequencies in both languages must be realized through the distributional patterns and the different conceptual tasks presented by nouns and verbs.

It seems quite plausible given the similarity of the distributional patterns of nouns and verbs in the two languages that Mandarin-speaking children as well as English-speaking children are rapid learners of object names. Children learning both languages may require minimal experiences with any individual noun category in order to learn it. And, more critically, children learning Mandarin and children learning English may learn nouns in the very same way. Mandarin children may produce fewer nouns simply because they hear fewer nouns. Analogously, Mandarin-speaking children like English-speaking children may work out individual verb meanings much more slowly than they work out individual noun meanings. Mandarin-speaking children may produce more verbs than English-speaking children, however, because they hear more repetitions of each verb and also more verb types. Type-token differences across different languages may not in-and-of-themselves significantly alter the nature of either noun or verb learning. This line of reasoning suggests we need to ask different questions about noun and verb learning in the two languages. For example, does the increased frequency of verbs in salient positions in the input to Mandarin-speaking children result in a different conceptualization of objects and events (see Kobayashi, 1997 for discussions on Japanese). At equivalent levels of vocabulary development do children learning Mandarin understand verb semantics better than English-speaking children but the simpler semantics of object names just as well?

Natural partitions hypothesis

Gentner proposed the natural partitions hypothesis in response to two sets of data: (1) the primacy of nouns over verbs in early vocabularies and (2) the universality of lexicalized noun meanings across languages but the high variability in lexicalized verb meanings. In the light of this evidence, she proposed that nouns are the starting point of word learning, early and uniformly because they are perceptually and conceptually straightforward. Recent evidence suggests that Gentner may have been wrong about the dominance of nouns in all early vocabularies and she might be wrong about nouns being the privileged starting point for all learners (Tardif, 1996; Choi, 1998). However, the natural partitions hypothesis may be right in the fundamental psychological divide that it proposes between nouns and verbs in early word learning.

REFERENCES

Anglin, J. M. (1977). *Word, object, and conceptual development*. New York: Norton.

Bloom, P. (1996). Controversies in language acquisition: word learning and the part of speech. In R. Gelman & T. K. F. Au (eds), *Perceptual and cognitive development*. San Diego CA: Academic Press.

Brown, R. (1973). *A first language: the early stages*. Cambridge MA: Harvard.

Carey S. (1978). The child as a word learner. In M. Halle, J. Bresnan & G. Miller (eds), Linguistic theory and psychological reality. Cambridge, MA: MIT Press.

Carey, S. (1982). Semantic development: state of the art. In E. Wanner & L. R. Gleitman (eds), Language acquisition: the state of the art. Cambridge: C.U.P.

Choi, S. (1998). Verbs in early lexical and syntactic development in Korean. *Linguistics* **36**, 755–80.

Choi, S. & Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language* **22**, 497–529.

Dale, P., Bates, E., Reznick, S. & Morisset, C. (1989). The validity of a parent report instrument. *Journal of Child Language* **16**, 239–49.

Fenson, L., Dale, P., Reznick, J. S., Bates, E., Thal, D. & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* **59**(5), No. 242.

Gasser, M. & Smith, L. B. (1998). Learning nouns and adjectives: a connectionist account. *Language and Cognitive Processes* **13**, 269–306.

Gentner, D. (1982). Why nouns are learned before verbs: linguistic relativity vs. natural partitioning. In S. A. Kuczaj (ed.) *Language development*. Vol 2: *Language, thought and culture*. Hillsdale, NJ: Erlbaum.

Gentner, D. & Boroditsky, L. (in press). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (eds), *Language acquisition and conceptual development*. Cambridge: C.U.P.

Gleitman, L. (1993) The structural sources of verb meanings. In P. Bloom (ed.), *Language Acquisition*. Cambridge MA: MIT Press.

Goldberg, A. E. (1998). The emergence of the semantics of argument structure constructions. In B. MacWhinney (ed.) *Emergenist approaches to language*. Mahwah, NJ: Erlbaum.

Goldfield, B. (1993). Noun bias in maternal speech to one-year-olds. *Journal of Child Language* **20**, 85–100.

Golnikoff, R. M., Mervis, C. B., and Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. . *Journal of Child Language* **21**, 125–55.

Higginson, R. (1985). *Fixing assimilation in language acquisition*. Unpublished doctoral dissertation, Washington State University.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology* **27**, 236–48.

Johansson, S. & Hofland, K. (1989) *Frequency analysis of English vocabulary and grammar: based on the LOB corpus*. New York: Oxford.

Kobayashi, H. (1997). The role of actions in making inferences about the shape and material of solid objects among Japanese 2 year-old children. *Cognition* **63**, 251–69.

Kuczaj S. (1977). *-ing, -s and -ed: a study of the acquisition of certain verb inflections*. Unpublished doctoral dissertation, University of Minnesota.

Lieven, E. V. M. (1978). Conversations between mothers and young children: individual differences and their possible implications for the study of language learning. In N. Waterson & C. Snow (eds), *The development of communication*. New York: John Wiley.

Macnamara, J. (1982). *Names for things: a study of human learning*. Cambridge MA: MIT Press.

MacWhinney, B. & Snow, C. (1990). The Child Language Data Exchange System: an update. *Journal of Child Language* **17**, 457–72.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge MA: MIT Press.

Markman, E. M. & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: taxonomic versus thematic relations. *Cognitive Psychology* **16**, 1–27.

Naigles, L. R. & Hoff-Ginsberg, E. (1995). Input to verb learning: evidence for the plausibility of syntactiv bootstrapping. *Developmental Psychology* **31**, 827–37.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* **38**(1–2, Serial No. 149).

Richards, B. J. (1994). Child-directed speech and influences on language acquisition: methodology and interpretation. In C. Gallaway & B. J. Richards (eds), *Input and interaction in language acquisition*. Cambridge: C.U.P.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology General* **104**, 192–233.

Sachs J. (1983). Talking about there and then: the emergence of displaced reference in parent–child discourse. In K. E. Nelson (ed.), *Children's language*, Vol 4. Hillsdale, NJ: Erlbaum.

Samuelson, L. K. & Smith, L. B. (1999). *Early noun vocabularies: do ontology*, *category structure and syntax correspond? Cognition*. **73**, 1–33

Smiley, P. & Huttenlocher, J. (1995). Conceptual development and the child's early words for events, objects and persons. In M. Tomasello & W. Merriman (eds), *Beyond names for things*. Hillsdale, NJ.

Smith, L. B. (in press). Learning how to learn words: an associative crane. In R. Golnikoff & K. Hirsh-Pasek (eds), *Breaking the word barrier: what does it take?* Cambridge: Oxford University Press.

Smith, L. B., Jones, S. S., Landau, B., Samuelson, L. K., Gershkoff-Stowe, L. (1999). *Origins of the shape bias*. Manuscript submitted for publication.

Snow, C. E. (1989). Imitativeness: a trait or skill? In G. Speidel & K. Nelson (eds), *The many faces of imitation*. New York: Reidel.

Stern, W. (1924). *Psychology of early childhood up to the sixth year of age*. New York: Holt.

Tardif, T. (1993). *Adult–to–child speech and language acquisition in Mandarin Chinese*. Unpublished doctoral dissertation, Yale University, New Haven , CT.

Tardif, T. (1996). Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. *Developmental Psychology* **32**, 492–504.

Tardif, T. Shatz, M. & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian, and Mandarin. *Journal of Child Language* **24**, 535 –65.

Tomasello, M. (1992). *First verbs*. Cambridge: C.U.P.

Waxman, S. R. (1994). The development of an appreciation of specific linkages between linguistic and conceptual organization. *Lingua* **92**, 229–57.

APPENDIX A

TRANSCRIPTS ANALYSED IN THE ENGLISH DATASET

Group	Child	Age	Source
< 1;3	May 01	0;11	Higginson (1985)
< 1;3	Andrew	1;2	Dale *et al*. (1989) & Snow (1989)
< 1;3	Margaret	1;2	Dale *et al*. (1989) & Snow (1989)
< 1;3	Naomi 01	1;2	Sachs (1983)
< 1;3	June 01	1;3	Higginson (1985)
1;4–1;8	June 03	1;4	Higginson (1985)
1;4–1;8	June 05	1;5	Higginson (1985)
1;4–1;8	Naomi 02	1;6	Sachs (1983)
1;4–1;8	June 09	1;7	Higginson (1985)
1;4–1;8	Naomi 03	1;8	Sachs (1983)
1;9–2;1	June 14	1;9	Higginson (1985)
1;9–2;1	April 01	1;10	Higginson (1985)
1;9–2;1	Naomi 27	1;11	Sachs (1983)

APPENDIX A (*cont.*)

Group	Child	Age	Source
1;9–2;1	Naomi 38	2;0	Sachs (1983)
1;9–2;1	April 02	2;1	Higginson (1985)
2;2–2;6	Naomi 51	2;2	Sachs (1983)
2;2–2;6	Naomi 53	2;3	Sachs (1983)
2;2–2;6	Naomi 60	2;4	Sachs (1983)
2;2–2;6	Sarah 13	2;5	Brown (1973)
2;2–2;6	Sarah 16	2;6	Brown (1973)
2;7–2;11	Abe 022	2;7	Kuczaj (1977)
2;7–2;11	Naomi 69	2;8	Sachs (1983)
2;7–2;11	Abe 038	2;9	Kuczaj (1977)
2;7–2;11	April 05	2;10	Higginson (1985)
2;7–2;11	April 06	2;11	Higginson (1985)

APPENDIX B

THE 50 MOST FREQUENT NOUNS, VERBS AND DESCRIPTORS
PRODUCED BY ENGLISH-SPEAKING CAREGIVERS LISTED IN ORDER
OF FREQUENCY

Noun	Verb	Descriptor
child's name	go	good
baby	want	little
book	put	pretty
dog	say	funny
kitty	look	nice
pig	get	green
ball	see	big
cat	like	blue
duck	think	broken
girl	know	yellow
juice	come	purple
hat	eat	new
bunny	got	kind
apple	read	red
house	thank	long
bird	play	careful
daddy	take	white
cow	remember	silly
shoe	try	old
eye	sit	wrong
cookie	tell	next
camera	make	hard
toy	talk	dirty
bug	ride	poor
box	hold	happy
hair	give	first
fish	bite	yucky
smurf	cry	tired

583

APPENDIX B (*cont.*)

Noun	Verb	Descriptor
donkey	turn	sandy
horse	tickle	black
hand	draw	brown
mommy	pull	great
sheep	stand	pink
ear	blow	bad
monkey	drink	gross
thing	fix	last
blanket	hurt	excited
chair	need	favorite
cheese	show	hairy
foot	move	sick
bear	sing	alone
car	find	afraid
face	hear	easy
raisin	help	full
water	made	hot
sock	understand	round
party	dance	salty
boy	use	sticky
paper	found	hot
doll	jump	wonderful

APPENDIX C

THE 50 MOST FREQUENT NOUNS, VERBS AND DESCRIPTORS PRODUCED BY MANDARIN-SPEAKING CAREGIVERS LISTED IN ORDER OF FREQUENCY

Noun	Verb	Descriptor
child's name	lai2 (come/arrive)	hao3 (good/ok/right)
a1yi2 (auntie/generic female)	kan4 (look)	xiao3 (little/small)
ma1ma (mommy)	chi1 (eat)	da4 (big)
che1 (vehicle)	shuo1 (say/speak)	xing2 (okay)
ba4ba (daddy)	deng3 (wait)	dui4 (correct/right)
shou3 (hand)	na2 (grasp/take)	huai4 (bad/broken)
fan4 (rice/generic food)	zou3 (go)	lao3 (old)
hua4(speech/condition/ drawing)	zuo4 (do/make/sit)	zang1 (dirty)
qiu2 (ball)	hua4 (draw/paint/cut)	duo1 (many/much/ a lot)
ren2 (person)	gei3 (give)	guai1 (well behaved)
jia1 (home)	jiao4 (call/make sound/ tell to do something)	chou4 (stinky/smelly)
zui3 (mouth)	wanr2 (play for fun)	hong2 (red)
shui3 (water)	da3 (hit/call/fire)	sui4 (broken)
nai3nai (paternal grandmother)	ting1 (listen)	teng2 (hurt/painful)

APPENDIX C (*cont.*)

Noun	Verb	Descriptor
hai2zi (child)	xie3 (write)	tao2qi4 (mischievous)
bi3 (pen/pencil)	he1 (drink)	hao3wanr2 (fun)
ye2ye (paternal grandfather)	dong4 (touch/move)	bai2 (white)
men2 (door/gate)	zhao3 (look for)	wan2 (finished/done for)
ge1ge (older brother)	shang4 (go to/ascend)	leng3 (cold)
mei4mei (younger sister)	fang4 (put)	hao3chi1 (yummy/ good to eat)
gu1 (paternal aunt)	qi2 (ride astride/start)	bang4 (terrific)
qi4che1 (automobile)	kai1 (open/drive/boil/is open)	hei1 (black/dark)
dong1xi1 (thing)	gan4 (do/work)	hao3kan4 (pretty/ good looking)
mao1 (cat)	guo4-lai2 (come over)	huang2 (yellow)
yu2 (fish)	nong4 (do/make/fix)	lei4 (tired)
jie3jie (older sister)	qi3-lai2 (rise)	yi2yang4 (same)
shir4 (matter/problem)	zhi1dao4 (know)	sheng1qi4 (angry)
shu1 (book)	ge1 (put/cut)	gao1 (tall)
ma3 (horse)	bao4 (hold/carry/hug/ register)	tao3yan4 (despicable)
deng1 (light)	chu1 (exit/emit/get out)	chen2 (heavy)
di4 (ground/younger brother)	shuai1 (gall/slip/throw)	ba3ba3 (yucky/dirty)
niao3 (bird)	xiang3 (think/desire/ miss/make sound)	e4 (hungry)
jiao3 (foot)	shang4 (sing)	gou4 (enough)
ya1zi (duck)	ku1 (cry)	re4 (hot/warm)
tang2 (candy/sugar)	gao4su4 (tell)	ke3 (thirsty)
tou2 (head/end)	tiao4 (jump/skip/dance)	chang2 (long)
wu1 (room)	zhan4 (stand/dip in)	gao1xing4 (happy)
hua1 (flower)	dai4 (wear/lead/take care of)	liang2 (cool)
dian4hua4 (telephone)	yao3 (bite/spoon out)	shi1 (wet)
pi4gu (buttocks)	la1 (pull/have a bowel movement)	luan4 (disordered)
xie2 (show)	xia4 (get down/scare)	tian2 (sweet)
qiang1 (gun)	jie1 (meet/receive/ continue)	piao4liang4 (pretty/ beautiful)
wa2wa (baby/doll)	mai3 (buy)	tang4 (hot/scalding)
pi2 (peel/skin)	reng1 (discard/throw)	la4 (hot/spicy)
bao1 (bag/blister/ swelling)	hui2-lai2 (go back near)	lan2 (blue)
shu1shu (paternal uncle/generic male)	pao3 (run)	po4 (broken/wounded)
lou2 (building)	ca1 (wipe)	xiang1 (fragrant)
huo3che1 (train)	chuan1 (wear)	ben4 (stupid/dumb)
gou3 (dog)	xi3 (wash)	yang3 (itchy)
deng4zi (stool)	diao4 (drop/turn around)	feng1 (crazy)