

Running head: Cognitive lexicon of the young child

Title: Beyond the realm of noun and verb: the cognitive lexicon of the young child

Christophe PARISSÉ and Marie-Thérèse LE NORMAND

Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France

Acknowledgments

This work was supported by a grant from INSERM, France: Contrat de Recherche Inserm (4U009B). Many thanks to Anne Reymond for her great help with the English version of this text.

Corresponding author:
Christophe PARISSÉ

Laboratoire de neuropsychopathologie du langage et de la cognition
INSERM - CRI 9609
Batiment Pharmacie, 3ème étage
Hôpital de la Salpêtrière
47 Bd de l'Hôpital
75651 PARIS cedex 13
FRANCE

Tel: 33 1 42 16 24 84
E-mail: parisse@ext.jussieu.fr

Title: Beyond the realm of noun and verb: cognitive lexicon of the young child

Abstract

Most studies of early lexical development are focused on the acquisition of the noun or verb categories. Only studies targeting the very beginning of word production describe the rich pattern of reference and expressive words produced by very young children. Still, during their second year, children's production in tokens contains as many words that are not nouns and verbs than words that are. The importance of categories such as communicators, adverbs, pointers and adjectives never decreases, neither in English nor in French children between the age of 1;6 to 2;6. A cross-linguistic comparison shows that the same type of words is the most frequent in English and French children, while a comparison with adult production shows that, in neither language, do the words produced by children match exactly the words they hear most frequently. The difference in the syntactic structure of English and French argues strongly for a cognitive origin to this close match of the children's words.

These words other than nouns and verbs are more complex than they appear, because they cover a whole range of reference principles – direct reference, indirect reference, shared reference, generic reference, multiple reference, ambiguity, similarity, repetition, absence of –, as well as a wide range of expressive meanings. This type of words appears and grows throughout the children's second year and provides the basic stones for further lexicon and syntax development.

INTRODUCTION

The nature of the first words produced and understood by young children is one of the oldest elements recorded in language development studies. These words are not created *ex nihilo*. Parents try often unconsciously to meet the child halfway, using motherese, gestures and generally behaving in a way adapted to exchanges with an infant. Conversely, children already have some knowledge of the world and a certain mastery of the phonetic structure of their mother language. From now on, they will have to create and master semantic indices and structures, and bridge the gap between the words as physical patterns, the meaning of their world and their interactions with it. Utterances will have to be analyzed and their constituents mapped onto the world constituents.

The process which starts at the end of the first year – when children produce their first words – and goes on to the end of the second year – when children begin to exhibit a real mastery of syntax – is a long one. Most studies of the children's second year are centered on the acquisition of the meaning of names and are focused either on the beginning of the single-word stage or on the period of the vocabulary spurt (Dromi, 1987). Studies about verbs usually start later, from around the age of two, at an age when verbs are often used in multi-word utterances. Very few studies are concerned with onomatopoeia, isolated deictics, adverbs, or adjectives during the single-word stage. Studies about these words are only done at the beginning of syntax, through their use in combination with nouns and verbs. The apparent reason for this is that children produce a high proportion of nouns during their second year. Gentner (1982) presents data that prove that nouns form the most important part of the first words learned by children. Fenson et al. (1994) give the following figures for words produced by at least 50% of children at 22 months: 63.2% of nouns, 8.5% of verbs, 5.3% of references to people, 7.4% of game and routines, 4.2% of adjectives, 5.8% of sound effects and 5.3% for others. The 'other' category includes words about time, pronouns, question words, prepositions, locations, articles, quantifiers, helping verbs and connecting words. The small proportion of 'others' is probably also due to the fact that these statistics can only be performed for types and not for tokens. A token analysis would have given a different result. Using a slightly different set of categories with the 24-month-olds of the Wells corpus in the CHILDES database (MacWhinney, 1995; MacWhinney & Snow, 1985), the results are: nouns 29%, verbs 20%, deictics and exclamations 16%, interjections 18%, adjectives 5%, negation 10% and questions 1%. This reveals a pattern more complex than the simple dichotomy between noun and verb. Significantly, the words belonging to the 'other' categories are also used in isolation, which adds up quite a lot of single-word utterances of semantic and pragmatic relevance. If the first computation is run through again, but for single-utterances only, the results turn out even more strikingly in favor of the non-nouns-and-verbs (non-N&V) categories: nouns 19%, verbs 11%, deictics and exclamations 19%, interjections 28%, adjectives 1%, negation 18% and questions 1%.

The small number of non-N&V types makes it all the easier for children to learn them, with the exception of the numerous exclamations which are often specific to each child. The use of these categories is very different from that of nouns and verbs, as it is both more general and more specific. It is more general because it refers to lot of varied objects or situations and more specific because it deals only with certain precise pragmatic circumstances, such as pointing, reference or the expression of a psychological state. Nevertheless, the high frequency of these words makes it impossible to pretend to understand the development of language without integrating them into the whole language development process. Because learning the multiple and ever changing interpretations of deictics and expressive words can be more complex than learning nouns, any explanation of the

acquisition of word meanings should take all the facets of child cognitive development into account – language, perception, motor control and affect – and not only the problem of reference, which in turn is not limited to nouns and verbs.

Single-word utterances in the literature

Works about the production of single-word utterances are numerous in the literature. This does not come as a surprise, as single-word utterances represent the very beginning of language in children and are easy to record and analyze as production is sparse. There are two very different types of works about the subject: those that deal with the very beginning of word production, usually before fifteen months, and those that deal with early lexical development, usually after eighteen months and up to the age of three. None of the first type of works (e.g.: L. Bloom, 1973; Dore, 1985; Griffiths, 1986; Griffiths, 1985; Werner & Kaplan, 1963) deals specifically with the learning of nouns or verbs, as they usually are not the first words produced by children, with the exception of the names referring to the caregivers. In fact, most first words are not standard words of the language, and are thus impossible to classify easily. Moreover, their exact reference is often holistic and difficult to pinpoint. As the first months of language production go by, the words uttered by children become more and more precise, in their phonetics as well as in their semantics. By the age of eighteen months, children already produce an average number of 170 words (Bates et al., 1994), with much possible variation from one child to another – from 3 to 544 words. At that age, children start acquiring words at a rapid rate (Woodward, Markman, & Fitzsimmons, 1994), and go through what is called the vocabulary spurt, which starts the beginning of a very long period of continuous and rapid vocabulary acquisition (Dromi, 1987; Dromi, 1999; Kuczaj, 1999; Reznick & Goldfield, 1992). The children's capacities for acquiring word and the great size and diversity of their lexicon have induced a quite different set of studies, aiming at understanding the principles underlying lexical development. Because nouns form the best represented category of words uttered by children at that age (Gentner, 1982), most studies deal only with the acquisition of nouns (P. Bloom, 1999; Gentner, 1982; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Huttenlocher & Smiley, 1987; Kuczaj, 1999; Macnamara, 1972; Macnamara, 1984; Markman, 1994; McShane, 1979). A typical example in this trend is Griffiths' (1986) which starts to describe the acquisition of the very first word before eighteen months within a framework involving all aspects of the child's development and then goes on to describe the sole acquisition of nominals at a more advanced age.

Authors have argued against this noun bias and some have shown that this noun bias is cultural and language dependent. In languages such as Korean and Chinese, verbs are learned at least at the same rate as nouns (Choi & Gopnik, 1993; Choi & Gopnik, 1995; Gopnik & Choi, 1995; Tardif, 1996). Studies such as Clark's (1993) deal with nouns acquisition as well as verbs acquisition. However, the problem is not just to substitute the noun bias with a verb bias. Other authors have argued against the very idea of a bias (L. Bloom, 2000; L. Bloom, Tinker, & Margulis, 1993; Nelson, Hampson, & Shaw, 1993). These critiques are very important because, as shown in L. Bloom (2000), the sole consideration of nouns in studies of lexical development is not without consequences on the theories developed. Word acquisition should not be isolated from the rest of the language development. L. Bloom (1993, 2000) makes a case for an intentionality model of word acquisition (see also Bretherton, 1988), a model that focuses on why children learn words and not only how, and which is based on principles such as relevance, discrepancy and elaboration, which are relevant to the child's motivation. In turn, this leads to the social implications in language development, and those works (Baldwin, 1993; Ninio, 1992; Ninio, 1993; Repacholi & Gopnik, 1997) that show the importance of joint reference between child and caregiver. The child's pattern of action is very important, as words are best learned in a situation where the objects denoted are used

(Dunham, Dunham, & Curwin, 1993). A simple description is often not enough when learning a new word.

In contrast, Markman (1989; 1994) describes principles such as the whole-object assumption, the taxonomic assumption and the mutual exclusivity assumption, which reduce the number of hypothesis a child is confronted to when learning a word. These principles do not describe the children's intentions or driving goals, whether implicit or explicit, in acquiring a language, but only give an external description of the processes involved. Golinkoff et al. (1994) make a proposal based on similar arguments with a theory that includes two levels of three principles each. This two-tiered approach takes the development of the children capacities during their second year into account. Some developmental processes are at work here, especially as the principles have to be overridden in some situations (Kobayashi, 1998; Landau, Smith, & Jones, 1998). This explains why the principles are also often called biases.

Goals of the study

The present article presents a study of the production of English and French children from age 1;6 to age 2;6. First, statistics of lexical production for every lexical category will be presented as well as a cross-linguistic comparison between English and French children. The comparison will bear on categories and lexical elements. Second, the same comparison will be extended to child vs. adult language, with adult cross-linguistic reference.

FIRST EXPERIMENT

Procedure

In order to prove that the children's vocabularies are complex and contain various and well-represented categories of words, it is necessary to use a typology of the children's lexicon. For this study, the typology will be the categorization of words in part of speech. This choice might look strange because there obviously cannot be any syntax when utterances are composed of single-words only. This does not in any way mean that children are sensitive to syntactic categories. Using syntactic tags is only a means of describing the child's language. Other choices are possible, especially using more semantic directed categories that are coherent with the children's production. Nevertheless, this particular choice has been made – and discussed – in numerous other studies (see for example Camaioni & Longobardi, 1995; Griffiths, 1986). The choice of part of speech is justified in that it is efficient in analyzing big amounts of data automatically, and provides a link with studies of more advanced language, using the same data and the same tools. Also, this approach is more neutral, with reference to the interpretation of children's language by adults and by researchers. There is no theory of children language development behind part of speech use in this article, only the choice of a typology tool.

All the material used in this study was tagged for part-of-speech using POST (Part Of Speech Tagger; Parisse & Le Normand, in press-a), an automatic tool integrated in the CLAN program (MacWhinney, 1995), which works for English as well as for French. Some slight modifications to the tagging scheme used in Parisse and Le Normand (in press-a) have been made in order to improve the ways to compare the English and French language. For the current article, the words yes and no will always be considered as adverbs, not as communicators. This offers a strict correspondence to the French tagging scheme and is grammatically correct (Random House Inc., 1997). No difference will be made between verbs and auxiliaries. This is once again grammatically correct, but would be a major mistake were the study of syntax the goal of this article or if older children were studied. As the current

work seeks to prove that nouns and verbs are not as important as is usually believed, mixing lexical and non-lexical verbs will merely add to the number of tokens in the verb category, which will not change the results, but on the contrary might even lessen their significance.

A lexical classification was then performed for each child, in all corpora, first on all utterances, secondly on single-words utterances only. Both type of analyses are useful as the study covers ages ranging from 1;2 to 2;6. Before the age of 2;0, single-word utterances represent nearly the whole of the children's production and, as they are the first words produced, they match more closely the beginning of language development than the few multi-word utterances produced. After the age of 2;0, single-word utterances evolve towards a different lexical structure and do not represent the child's language correctly any more. Comparing the two sets of words offers useful insight into the development process. Unless stated otherwise, all statistics are performed across transcripts.

Material used for the English children

The English language material comes from the CHILDES database. Three corpora have been used for the present experiment. All children described in these corpora have a normal language development. The first corpus contains transcripts from the Bates/Bretherton Colorado longitudinal sample of middle-class children studied in Boulder, Colorado, between 1978 and 1980 (Bates, Bretherton, & Snyder, 1988; Carlson-Luden, 1979). Only the free play interactions have been used for the current work. All free play sessions for the same age group used the same instructions and the same toys. The second corpus was gathered by Catherine Snow, Barbara Pan, and colleagues, as part of the project 'Foundations for Language Assessment in Spontaneous Speech', project funded by the National Institutes of Health, USA, (Ninio, Snow, Pan, & Rollins, 1994). Transcripts reflect spontaneous language data collected during a five-minute warm up and several subsequent activities, which include playing with toys, mother-child interaction and some mother-child-examiner interactions. The third corpus comes from a project by Gordon Wells and colleagues: "The Bristol language development study: language development in pre-school children's" (Wells, 1981). Each observation consists of a recording in the child's home and tests at the Research Unit of the university. Spontaneous conversation forms the main part of each observation. Only those transcripts with children aged less than 2;6 have been used for the current study. The characteristics of all the English transcripts used in this study are given in Table 1. The resulting corpora make a total of 37,590 words.

insert Table 1 about here

Material used for the French children

insert Table 2 about here

A systematic evaluation of the development of lexical categories in young children has been done using a database (see Table 2) created following a technique of direct observation of behavior samples (Le Normand, 1986). It uses direct spontaneous speech data produced during symbolic play, in the same standard situation, video-recorded openly, always by the same observer. The recordings were made in a play situation to let the children comment on their own actions, talk about real or imaginary events, and converse with a familiar adult partner. The strictly standardized play material consists of a family house with five figurines, eleven pieces of furniture (two tables, four chairs, two armchairs and three beds), and five figurative objects (stairs with a mobile door, garage with a sliding door and a front door bell).

For the data gathering, the technique of full sampling of behaviors was used. Child speech was segmented into utterances using the criteria defined by Rondal, Bachelet, & Pérée (1985), in accordance with the CHAT system (MacWhinney, 1995). The corpora presented here range from the age of 1;8 to the age of 2;6. The children have a normal pattern of linguistic development and are raised in a monolingual environment. The characteristics of the corpus are listed in Table 2. The full corpora make a total of 17,919 words.

RESULTS AND DISCUSSION FOR THE FIRST EXPERIMENT

Use of lexical categories by the English children

insert Table 3 about here

Table 3 presents the English children's lexical characteristics for all utterances, with scores averaged separately for each corpus. Three types of results are presented for each corpus. First, the percentage of words belonging or not belonging to the noun-and-verb (N&V) category, either not. Second, the percentage of words belonging to one of the following categories: noun, verb, adjective, adverb, communicator, wh-word, pointer – i.e. demonstrative pronoun, existence pronoun and possessive pronoun – and the 'other' category for all words not in the previous categories. These categories were chosen because they cover most of the words produced in isolation by children. Consequently, the 'other' category represents mostly words that are syntactic markers or produced in the context of other words. Third, the percentage of children that produced at least one word for a given category.

The first type of results shows that, in tokens, there are roughly as many words that are nouns or verbs than words that are not. This is true at every age, with the exception of the very young children of the Snow et al. corpus and all the children of the Bates et al. corpus. In this first exceptional case, nouns and verbs make up only 24% of the words and the other words are mostly communicators. These children are very young and barely speaking yet, so most of their production still is onomatopoeia. In the second case, nouns and verbs make up only a third of the words. The difference between, on the one hand, the Bates et al. corpus and on the other hand, the Wells corpus and the older children of the Snow et al. corpus is difficult to explain because the children's age and the recording situation are apparently similar.

The second type of results shows that the communicator category is as much represented as the categories of nouns and verbs. The importance of this category tends to decrease with age, but remains high after age two. The categories of adverb, adjective, wh-word and pointer are nearly equally represented with an advantage to the pointer category – average percentage across age are 2.1, 3.3, 1.8 and 4.5, respectively. These values may appear quite low in comparison with the values for noun, verbs and communicators. However, the real import of these categories for young children is better shown by the third type of results below. More fine-grained categories may explain the differences found above between the Bates et al. corpus and the other corpora. The Bates et al. corpus contains only play situations – which is not the case for the two other corpora – and this may elicit less naming of objects and more naming of actions and communicators. Indeed, the percentage of nouns is much lower for the Bates et al. corpus than for other corpora and the percentage for communicators higher. In addition, the transcripts of the Bates et al. corpus contain fewer tokens, which may indicate that children were more focused on playing than on describing the surrounding environment.

The third type of results reveals how many children use a category, whether rarely or frequently used. It helps to know if a category that has a low frequency is nonetheless used quite systematically across all ages and corpora. The first interesting result is that not all of

the very young children of Snow et al. produce nouns or verbs. Only 53% produce nouns and 27% verbs. This does not come as a surprise because most of these children speak very little and many of them have started to produce only a few onomatopoeia and not very many real adult words. The second interesting result is that categories that represent only 1% in tokens can be used by more than a third of the children. Categories that represent 2% can be used by more than half of the children. This is true even for children aged 1;6 or 1;8, an age where the total number of tokens is quite low. For these children, 2% of tokens represents only two tokens. For older children, 2% represents up to six words. In these cases, a category which represents 2% of the tokens is used by 74% of the children. This is true for the adverbs of the Wells children at age 2;0 and the wh-words of the Wells children at age 2;6. This also explains why the percentages for the Bates et al. corpus at age 1;8 are low. The average number of tokens per transcript for this corpus is 29, so that one single token makes up to 3% of the total number of tokens for a transcript. Non-frequent categories are clearly at a disadvantage in this corpus. For the rest of the corpora and for the adverb, adjective, wh-word and pointer categories, an average of nearly 50% of the younger children produce these categories. At age 2;0, an average of 75% of the children produce these categories. Adverbs tend to appear a bit later than adjectives and pointers, and wh-words appear last of all.

insert Table 4 about here

The classification into the categories of noun, verb, adjective, adverb, communicator, wh-word, and pointer was justified on the basis that these categories made up the bulk of the words produced in single-word utterances. Their production in isolation attests that children use these words meaningfully, and if not with the usual adult meaning, nevertheless with an interpretation of their own. When used in multi-word utterances or frozen-forms, these words are good candidates for explaining what type of information the child use to build and understand speech.

All the variables presented in Table 3 have been computed again using single-word utterances only. Results are presented in Table 4. It demonstrates that all of the seven categories above are used in single-word utterances. Looking at the differences between the percentages of tokens per lexical category for all utterances and for single-word utterances, one can see that communicators are more common in single-word utterances than any other class. Nouns tend to be used a lot too, but no more than for all utterances. All other categories are used less in isolation than for all utterances. The differences are especially high for the 'other' categories. The higher the difference, the older the child.

Use of lexical categories by French children and comparison with the results for English children

The French data was tagged for part-of-speech in the same way that English data was, using the same tool, POST. The statistics presented for English in Table 3 are reproduced for French in Table 5, those of Table 4 reproduced in Table 6. The same set of categories was used with one exception: wh-words were replaced by relative pronouns, because this is the closest French category to be found, although the two are not exactly equivalent. For each age group, means were computed across children.

insert Table 5 about here

The general trend of results obtained for English holds for French (compare Table 3 and Table 5). The average percentage of the proportion of nouns and verbs is nearly the same in French (42%) and in English (43%). There are nonetheless some important differences. Looking at individual categories, the proportion of nouns, communicators and adjectives is

higher in English (respectively 31%, 24% and 5%) than in French (respectively 22%, 17% and 2%). Conversely, the proportion of verbs and pointers is lower in English (respectively 12% and 4%) than in French (respectively 20% and 15%). The difference between the English and the French children in their use of nouns, verbs, adjectives and pointers is significant (respectively, $t(12) = 2.43$, $p = .03$, $t(12) = 2.80$, $p = .02$, $t(12) = 3.44$, $p = .005$, $t(12) = 4.01$, $p = .002$). The proportions for the other categories are the same in French and in English.

In the percentages of children producing at least one word from a given category, English children produce more systematically nouns, communicators, adjectives and ‘others’, whereas French children produce more systematically pointers and relative pronouns. Significant difference is found only for pointers (respectively $t(12) = 2.67$, $p < .02$).

insert Table 6 about here

For single-word utterances (see Table 4 and Table 6), the results of the comparison made between English and French children are nearly identical to those obtained for all utterances. There are more isolated nouns in English (36%) than in French (25%) but fewer verbs (5% in English, 18% in French). Communicators are more numerous in English (33%) than in French (27%), and so are adjectives (5% in English, 1% in French), but the opposite holds for pointers (2% in English, 18% in French). Significant differences are found for the same categories: nouns ($t(12) = 2.31$, $p = .04$), verbs ($t(12) = 15.03$, $p < .00001$), adjectives ($t(12) = 3.73$, $p = .003$) and pointers ($t(12) = 5.10$, $p < .0003$).

The only difference is for the ‘other’ category, which is more frequent in English (5%) than in French (1%). The difference is significant, which was not the case for all utterances ($t(12) = 6.44$, $p < .00001$).

Lexical comparison between English and French children

insert Table 7 about here

A close analysis of the words that produced a trend towards non-N&V words is necessary to understand the function of these words. Table 7 shows the most frequent words for English and French, presented using the CHAT format (<http://chilides.psy.cmu.edu>; MacWhinney, 1995). In order to be able to compare ages where the number of words produced are very different, the tokens are not rated by their raw frequencies. Instead, the lists of the twenty most frequent words were first determined for each corpus, which made it possible to rate the words by the number of times they appeared in these lists. The best scores for each language are presented in Figure 7, for all utterances and for single-word utterances. If a word scores nine in English – for example the word oh –, this means that it appeared nine times as one of the twenty most frequent words of a corpus. Nine is the maximum score for English – there are nine English corpora – and five is the maximum score for French.

There are no words such as v:be&3S (corresponds to is and 's), det|the, det|a and pro|it in the most frequent words of the single-word utterances because, as they have no meaning by themselves, they are only produced with another word. The same holds for the French words v:aux|être&3S (corresponds to est = is in English), v:aux|avoir&3S (corresponds to a = have in English), det|le (the in English, but in the masculine singular only), det|la (the in English, but in the feminine singular only), adv:neg|pas (not in English), v:mdl|aller&3S (corresponds to va = is going to in English), pro:c' (it as subject pronoun in English), pro:subj|il (he in English or it when used as an impersonal pronoun) and pro:obj|l' (it as object pronoun, him or her in English). There are nine words of this type in the twenty most frequent words in French and only four in English. This confirms that, as demonstrated elsewhere (Bassano, Maillochon, & E., 1998; Parisse & Le Normand, 1997; Parisse & Le Normand, in press-b),

functional words are produced sooner in French than in English, although the most frequent functional words are quite similar in both languages. There is a great correspondence between the functional words produced by English and French children. The French equivalent of the four most frequent English words also figure among the most frequent French words.

insert Table 8 about here

Words can be classified into three broad categories: nouns and verbs, functional words and the rest of the words. Functional is not defined as the set of close-class words here, but as words that do not appear in isolation, such as those presented above. Words such as that, there or what, which could be considered as functional in some cases, appear in single-word utterances with a standalone meaning – see footnote 1. In these cases, they do not simply complement another word. Table 8 presents the distribution of the most frequent words into those three categories. Clearly, there are more words from the other categories in English than in French. There are more nouns and verbs than words from the other categories in English and French, although this is more pronounced for French. This looks a bit at odds with the overall trend to find less functional words than content words in young children's production in English. The contradiction holds only if communicators and pointers such as there are considered as functional words, which would go against the classification suggested above. In this classification, functional words are limited to words which have a morphological value and do not always appear as separate words in agglutinative or flexional languages – e. g. Turkish or Russian; the difference between nouns and communicators is smaller than between nouns and functional words because both participate in the cognitive world of the child with the similar syntactic value, 'isolated word'. The difference between English and French could come from the difference in the recording situations. In the English corpora, many recordings were done in a context of interaction between child and adult, thus emphasizing the communicator category, whereas in the French corpora, the recordings were done in a play situation that led children to describe their acts and the situation, thus emphasizing nouns and verbs.

In English, the fourteen words among the twenty most frequent single-word utterances for all ages which are not noun, verb or purely functional are all categorized as communicators except for three words: there, a demonstrative pronoun, existence pronoun or adverb of location, depending on the grammar, what, a wh-pronoun and me, an object pronoun. The communicators come into three types: standard interjections such as uh, oh, ah, huh, ow, non-standard interjection such as mm and um, and adverbs such as no, yes, yeh and yeah. In French, there are eleven such words, and among them one demonstrative pronoun ça (that), five adverbs and five communicators. Two of the adverbs could be classified as communicators, oui (yes) and non (no), two are adverb of place, voilà (there it is) and là (there), and the last adverb, encore, is the French equivalent of more in children's language – more does not figure in Table 7, but comes close in the 26th position for single-word utterances. French interjections are ah (ah or oh in English), oh (oh in English), dodo (bed), coucou (hello, peek-a-boo) and hein (eh, what). There is a high correspondence between English and French words: twelve of the most frequent English words have their closest translation in French placed amongst the twenty most frequent French words. When words do not appear in either lists, they either come not far after the 20th place, or else have a close equivalent in the lists. For example, ça corresponds to that which occurs in 26th position for single-word utterances and in second position for all utterances, moi (me) appears in 44th position for all utterances and in 171th position for single-word utterances, hello appears in 36th position and voilà is another equivalent to there. The correspondence also exists in nouns and verbs, when the three nouns and one verb most frequent in English can also be found amongst the most frequent French words – e.g. look = regarder, mummy, mum, or ma =

maman, daddy = papa and car = voiture. However, this is not true of the other French words in the list, v|tenir (tiens = take), v|attendre (attends = wait) and v:pp|fermer (fermé = closed). There is a close equivalent in English, gone, in the 37th position, for v:pp|partir (parti = gone).

SECOND EXPERIMENT

The first experiment showed that English and French children often produce words which are neither nouns nor verbs, especially in their single-word utterances. Are these words copies of the children's input, or are they constructions made by the children? The answer to this question is of importance as the former would mean that it results from a tendency to follow the patterns of language – a syntactic principle – and the latter that it results from cognitive processes – a semantic principle. It is thus necessary to select English and French adult corpora and process them in an identical fashion to the children corpora, so as to compare children and adult use of words.

Material used for the English adults

For the English, all corpora contain the transcriptions of the adult's utterances, so that the procedure applied to children's utterances can be repeated. It is possible to compute the distribution into syntactic categories for words occurring in all utterances and in single-word utterances, and to compute the list of the most frequent words. There is a total of 316 adults throughout the corpora, with an average MLU in words of 3.56 (SD = 0.25), an average proportion of single-word utterances of 24% (SD = 2.7%) and an average number of words per transcript of 510 (SD = 300). The whole of the corpora makes a total of 377,745 words.

Material used for the French adults

As the Le Normand corpora do not contain any transcription of adult utterances, it was necessary to find other corpora to study French adult language. The first selection was the adult data gathered by Madeleine Léveillé with the participation of Patrick Suppes (Suppes, Smith and Léveillé, 1972; Suppes, Léveillé and Smith, 1974) and available in the CHILDES database. This corpus corresponds to thirty-three tape recordings of an hour each, of a child at home, covering the ages of 2;1 to 3;3. The transcripts include both the utterances of the child and of the adults, namely the mother and father of the child and field researcher Madeleine Léveillé. Every sentence was analyzed and the utterances divided into children and adult utterances. The adult part of the corpus, referred to hereafter as the 'Léveillé corpus', contains 19,891 adult utterances corresponding to 120,563 words, with a MLU in words of 6.03 (SD = 0.45 across transcripts).

The second selection was the adult data of the corpus contributed by Jean Rondal of the University of Liège (Rondal, 1985). It contains a set of 120 transcripts of interactions between Jean Rondal's son and his mother, covering the ages of 1;2 to 4;9. The transcription and segmentation of utterances were done in accord with the principles discussed in Rondal et al. (1985). The bulk of the adult data comes from the mother, but occasional sentences from the father or other adults are also included. The MLU in words is 5.18 (SD = 0.60 across transcripts). The full adult corpus makes a total of 238,438 words.

A third corpus is made of the transcripts of a play situation where children were asked to exhibit a standard behavior and the adult's role was to prompt the child for language if necessary. The transcripts correspond to different children aged from three to four and the recordings were done by students in our laboratory. The transcriptions were done in the same format as those of the Le Normand corpus of children above. The MLU in words is 7.14 (SD = 1.13 across transcripts). The adult part of the transcripts makes a total of 38,991 words.

The three adult corpora make a grand total of 397,992 words.

RESULTS AND DISCUSSION FOR THE SECOND EXPERIMENT

Lexical characteristics of the adult corpora in English and French

insert Table 9 about here

insert Table 10 about here

As there is not much variability between the different English adult's corpora, it is not necessary to present full tables describing each corpus individually. The lexical characteristics presented in Table 3 and 4 for the children were all computed for the adult's corpora. Average results and standard deviations for English adults are presented in Table 9.

The same remarks hold for French. The variations between the different French adults are small enough that no individual presentation of adult lexical output is necessary for the current work. Average results and standard deviation for French adults are presented in Table 10.

A short appraisal of the lexical category distribution for the English and French adults' corpora (see Table 9 and 10) and a quick comparison with the children's distributions shows similar trends in English and French. There are some major differences in the adult's lexical distribution for all utterances and for single-word utterances. There are fewer verbs produced in single-words utterances but more communicators and adverbs and the 'other' category is almost non-existent in single-word utterances. For the percentages of adults producing at least one word from a given lexical category – bottom of Table 9 and 10 –, all categories come close to 100% of production for all utterances, except for the single-word utterances, where categories such as nouns, verbs and communicators have a high production rate, but not a 100% rate. Adjectives, wh-words, pointers and 'other' have lower rates, especially in French. In English, adjectives and wh-words are produced in isolation just as much as verbs are, over than 70%.

insert Table 11 about here

Table 11 shows the most frequent words for English and French adult language, for all utterances and for single-word utterances. The figures in Table 11 correspond to percentages of occurrences for all corpora. They should not to be compared with the figures of Table 7 which were not percentages of occurrences, but the number of times the word appeared in the list of the twenty most frequent words. It is interesting to see that, as was the case for children, there is a certain degree of correspondence between the words used by English and by French adults. Nine English words in the twenty most frequent ones for all utterances find their natural and closest translation in the French corresponding list and, conversely, so do fourteen French find words their equivalent opposite (you = tu, is = est, the = le or la, it = c' or on, a = un, not = ne and pas, in = dans, what = que, and = et). For single-word utterances, the correspondence is higher, thirteen English words and fifteen French words have a correspondence (okay, yeah and yes = oui, oh = oh, ah, or ahah, no = non, look = regarde, what = quoi and hein, there = là, here = voilà, eh = hein, mm = mm, good = bon). Some of the equivalences are not exact because the mappings between form and function of the words or exclamations are not the same in the two languages (for example, oh in English and oh or ah in French, here in English and voilà in French).

Comparison between adult and child distribution of lexical categories

In order to compare child and adult distribution of lexical category production, it is possible to compute the Pearson product-moment correlation between the percentages of categories produced by children and by adults. The categories include all those defined above, noun, verb, communicators, adverbs, adjectives, wh-words, and pointers, as well as the 'other' category – the rest of the words. Correlations are computed, for each language, between all the adult corpora regrouped in one big corpus and, turn by turn, each children's corpus which by-passes the lack of similarity between English adult and French adult corpora.

insert Table 12 about here

Table 12 gives the correlation value for all children's corpora for both English and French. Results are computed separately for all utterances and for single-utterances. There is a clear increase in the correlation values for all utterances, starting from a total absence of correlation under age 2;0 up to a very high correlation at age 2;6. This is true for English and for French, with very similar values. But there is a major difference for single-utterances between English and French. In English, child single-word utterances correlate very well with adult's. This is true at any age and the correlation values remain the same throughout, whereas in French, the correlation is low at any age and remains so throughout.

Lexical comparison between children and adults

The results in Table 7 and Table 11 have been used to compare the most frequent words produced by children, in English and in French, then by adults, in English and in French. They can also be used to compare the most frequent words of children and adults, for English and for French. In English, out of the children's twenty most frequent words extracted from all the utterances, only twelve of them belong to the adult's twenty most frequent words. There are nine such words for single-word utterances. In French, the figures are eight twice. Most words that are not in the twenty first words can be found in the fifty most frequent words, for English and for French. Nonetheless, it remains true that English words such as oh, no, that, there, uh, yes, what and ah, and French words such as pas, là, voilà, oh, ça, ah and non are much more frequent in child language than in adult language.

insert Table 13 about here

Table 13 gives, for every English corpus, the full list of the twenty most frequent words produced by the children and the rank of each word in the adult lexicon of the corresponding corpus. The number of words which are neither nouns nor verbs – nor functional words – tends to remain constant through age. The number of frequent adult words produced by children tends to increase.

insert Table 14 about here

This can be seen in Table 14 which presents the average rank of the twenty most frequent words for English children and for English adults, corpus by corpus – with separate English adult corpora. The reference used to compute the rank value is the full list of the words produced by the adults of all English corpora, making it possible to compare the average rank of children and adults' words. There is an evolution in the average rank for children's words that follows the children's age, whereas there is no difference from one adult corpus to another.

insert Table 15 about here

insert Table 16 about here

There are two main results for the current study. The first is that there are – in tokens – as many or more words that are neither nouns nor verbs than words that are. This is true for English, which reproduces and confirms the results of Bloom et al. (1993), and for French. It applies, not only to the period before the age of eighteen months, but also to the rest of the children's second year, and to the first half of their third year. This is confirmed by the results presented in Table 15 and 16 where the number of word types steadily increases in all categories and at all ages (as the number of word types is a function of the length of the recordings, comparisons between different ages should be done only for similar corpora, e.g. Snow, Bates or Wells). It is not justified to study young children's linguistic competence on nouns and verbs only and ignore the other categories, thus ignoring a large part of what their production is. All word categories should be included, nouns and verbs as well as other categories. The second result is that the words used by children do not simply follow the frequencies of the children's input (see Table 13 and 14). The correlation between child production of lexical categories and the corresponding adult production is summarized in Table 17. Cross-language correlations are presented vertically, intra-language correlations horizontally. Results are presented for all utterances and for single-word utterances. The most important information given in Table 17 is that, for all utterances, the correlation between children at a given age and adult starts very low and increases steadily with age, whereas the correlation between all English and all French children is high (see footnote 2). The correlation effect between children using different languages is most probably due to a semantic effect, which is confirmed by the great lexical correspondence between English and French children, as shown in Table 7 above. The results in Table 17 for single-word utterances also show that there is no reproduction effect of the adult's single-word utterances. Taking only into account English data, it would be possible to be lead into believing that children are reproducing the format of their input, because of the very high correlation between adult and child distribution of single-word utterances. But this is not true for French at all. The French adult's system of single-word utterances is completely different from the English adults' and from the children's, whether English or French. On the contrary, the correlation between English and French children for single-word utterances production is high. As there is no possible syntactic principles at work in the child's production of isolated words and as the systems of syntactic categories used for English and French in this article are the same, the single-word utterance correlation can only have a semantic origin. As a consequence, it appears that children create and follow their own semantic agenda when they began to learn a language (Bloom, 2000). They are not like a sponge, automatically absorbing all that surrounds them, but rather interact with the world. This confirms the results of Baldwin (1993) and Tomasello & Kruger (1992), which show that joint reference and the child's own actions are at the root of word learning.

insert Table 17 about here

It should come as no surprise that children speak in their own words, even though these words do come from their language input. One clear example of this is the word car, which

French equivalent is also used a lot by French children. This is obviously a cultural choice of word, but the children's great interest for that word comes more from playing with a toy car and riding in the family car, than from hearing the word uttered by their caretakers. A similar example is the case of the personal pronouns, which was not addressed in the discussion as it is not the focus of the current article. Looking at Table 11, the most frequent personal pronoun produced by parents is you in English, and tu in French –its exact translation. The most frequent personal pronoun produced by English children is it – an impersonal pronoun. For French, it is the same: c', the most frequent pronoun, is also impersonal (see footnote 3). As there is no great difference in phonemic complexity between the various personal pronouns and also no difference in their grammatical complexity, the only possible causes of the different use of pronouns by children are semantics and pragmatics.

So, if children use their own words, what is the cognitive status of words such as oh, no, that, there, uh, yes, what, ah, and yeah? These words are the most frequent words produced by English children that are not pure functional words. Only one verb, look, and one noun, Mummy, are more frequent than yeah, though less so than ah. Interestingly, look is not an object word but a special type of action word, a relation word in fact. It is obvious that Mummy also has a very special status in the child's cognitive universe. The same remarks applies to French children with only one very frequent concrete word voiture (car), but this is probably a consequence of the sample recording procedure which puts the child in a more descriptive situation than, for example, a conversation.

Child words serve a large variety of semantic functions (Nelson, Rescorla, Gruendel, & Benedict, 1978). They also have other functions, expressive or pragmatic (Halliday, 1975; McShane, 1980). These two functions are often used to classify children's words (Barrett, 1986; Dore, 1975) and language behavior. Their interactions are not usually considered in the theories that ignore joint reference and the fact that children lead the interactions with adults. No word has an expressive or a reference function only. This is true for adults and for children. For example, to say hello means that one has identified that a person has just arrived and refers to this identification. It also implies many other things about the relation between the two persons involved. Bretherton (1988, p. 234) gives some examples about children's use of the word bye. In one case, the child was reported to use bye with handwaving, but out of context. In another case, a child used bye only after the person had left. Two other children seemed to comment an impending departure. Finally, another child used bye as a request to have the door opened. It is difficult to know exactly what was the expressive function of these occurrences from the child's point of view, but it is clear that children were in the process of learning what bye referred to. And from a logical point of view, this is much more complex than simply learning to associate a word with an object. Objects are much easier to isolate from the context. Learning what is referred to in the act of waving goodbye involves understanding what a person is, the fact that a person is there or not, the impending change of state involved, the difference between arriving and leaving, and the fact that this applies to any person. A child that has fully understood that will probably learn to name an object easily – and be older than two. This knowledge does not come up suddenly as a whole. It is acquired step by step, starting with some contextualized situations and people and gradually developing towards more generic knowledge, in a manner reminiscent of the verb island hypothesis of Tomasello (1992). Children also learn the complex reference system of many other words in their second year, which are not nouns, and these words turn out to be the most frequent words they use – see Table 7, with the exception of Mummy and Daddy which are nouns. And other frequent words such as more, mine, down, gone, up, etc. should be added to the list. A simple word such as oh has multiple possible meanings. It is usually used as an expression of surprise, pain, or disapprobation, but also to attract the attention of the person spoken to. Children often use it to comment on an action, either in order to share that

comment with an adult or simply when playing alone. But most of the time, it explicitly refers to an action. Using such a simple quasi ‘primitive’ word, children begin to learn what a generic reference is and what an action is. They continue to do so with no and with look. Words such as there and that are more targeted towards objects – still with a generic reference. If all these words are heavily associated with expressive functions, but also with referential functions, the opposite holds for objects words. Words such as cat, dog, bed, or eat, are used for expressive purposes as well as for reference purposes.

During their second year, children learn to differentiate all the uses, implications, references, and phonetics of the words they produce. This behavior has been described by Werner and Kaplan (1963) and L. Bloom (2000). The children’s attitude is governed by interaction with other people, which serves as a drive and a frame for language use. When they starts learning objects words quickly and steadily, children have already mastered a lot of complex linguistic mechanisms such as generic reference, indirect reference, reference to an object, reference to the use of an object, reference to a specific individual, and reference to unknown people. They already know that the same word can refer to more than one thing (for example, baby for themselves but also for other children). At the same time, their phonetics – production and comprehension – has become much better, and general behavioral capacities such as a full mastery of object permanency is achieved or on the way to being achieved. It seems reasonable to think that all those behaviors converge towards the mastery of the linguistic media, which will in term probably help other knowledge to develop. The children’s global linguistic and behavioral capacities enable them to handle problems such as those raised by Markman’s assumptions in an almost conscious and natural manner. One of the bad consequences of studying nouns only – because they are most common category produced – is to forget that children already have and exhibit complex capacities when using other words. This cognitive and linguistic knowledge has potential consequences, not only on single-word utterances processing, but also on multi-word utterances processing. The progressive differentiation in the use of words makes it easy and natural to begin to build utterances of more than one word. The same thing goes for comprehension. It is therefore logical that the beginning of multi-word utterances often closely follows the start of the vocabulary spurt, regardless any aspect of morphosyntactic knowledge.

What is missing in most descriptions of principles, assumptions, or biases of word acquisition is that no proposal is made about how these biases come to be. Theories that relies on biases only tend to be more descriptive than explicative. Theories such as those of L. Bloom (2000), Locke (1996), Tucker and Hirsh-Pasek (1993), and Werner and Kaplan (1963) are more directed towards explaining why and how children speak than describing child speech. Only these types of theories can make strong predictions and be easily tested. Whatever their underlying principles, all have in common that they do not dissociate a part of language from the rest of language, or language from other behaviors. Such dissociation is probably at the heart of a misunderstanding in the problem of object naming. The famous example of Quine (1960) is described in the context of a translation problem (see footnote 4). It is true that translators are often in a situation where they hesitate between different meanings and interpretations, but the problem of the child is very different. Children are not Quine’s linguist. They don’t know any language and their knowledge of the world is much different.

Children do not hesitate nor choose between different hypotheses. They are in a certain context of action or interaction, looking at a novel situation with their developing knowledge of the world and perceptive capacities. They are not in a hypotheses making situation. Either they are able to understand or get something specific from a situation, using their linguistic or world knowledge, and they will remember something specific, or – if children are indeed able to do this – they store the whole situation in an eidetic fashion and will use that knowledge

later in new circumstances and with any new knowledge available. Later, when hearing the same word, they will again be in a situation where some natural action can be performed and their understanding will become more detailed and closer to the adult's. No deduction nor hypotheses filtering needs to be done. No conscious attribution of a name to an object nor selection needs to be done. The salient elements of the language input will become associated with the salient elements of the situation, according to the understanding of the situation by the child. This explains why biases can be overridden because children do not understand every situation in the same way. Thus they do not always associate words and objects, persons or actions in the same way. Sometimes they will focus on the whole object, sometimes on a specific part of the object, depending on the circumstances. If children have a bias, it is to use language and interact with the world. It is true that sometimes children do not know immediately what to do and remain thinking, obviously hesitating. But it would a little hasty to say that they go through a hypothesis pruning process. If only because their language is not developed enough, their deduction process is most probably unconscious and non-verbal. Unconscious deduction in adults is something not really understood, but it is not usually described as a process of filtering hypotheses. On the contrary, it often seems as if the real problem were to find a hypothesis, any hypothesis. We don't know if this apply to children, but this stresses the fact that it is important to place situations in their context. The whole system composed of children's language, cognition, and physical capacities has to be kept in mind when interpreting their behavior.

REFERENCES

- Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. Journal of Child Language, 20(2), 395-418.
- Barrett, M. D. (1986). Early semantic representations and early word-usage. In S. A. Kuczaj & M. D. Barrett (Eds.), The development of word meaning . New York, NY: Springer-Verlag.
- Bassano, D., Maillochon, I., & E., E. (1998). Developmental changes and variability in the early lexicon a study of French children's naturalistic productions. Journal of Child Language, 25(3), 493-531.
- Bates, E., Bretherton, I., & Snyder, L. S. (1988). From first words to grammar. New York, NY: Cambridge University Press.
- Bates, E., Marchman, V., Thal, D., Lenson, L., Dale, P., Reznick, S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. Journal of Child Language, 21, 85-123.
- Bloom, L. (1973). One word at a time: The use of single word utterances before syntax. The Hague, The Netherlands: Mouton.
- Bloom, L. (1993). The transition from infancy to language. Cambridge: Cambridge University Press.
- Bloom, L. (2000). The intentionality model of word learning: How to learn a word, any word. In R. M. Golinkoff, K. Hirsh-Pasek, N. Akhtar, L. Bloom, G. Hollich, L. Smith, & M. Tomasello (Eds.), Becoming a word learner: A debate on lexical acquisition . Oxford, NY: Oxford University Press.
- Bloom, L. M., Tinker, E., & Margulis, C. (1993). The words children learn: Evidence against a noun bias in early vocabularies. Cognitive Development, 8(4), 431-450.
- Bloom, P. (1999). Theories of word learning: rationalist alternatives to associationism. In W. C. Ritchie & T. K. Bhatia (Eds.), Handbook of child language acquisition . San Diego: Academic Press.
- Bretherton, I. (1988). How to do things with one word: the ontogenesis of intentional message making in infancy. In M. D. Smith & J. L. Locke (Eds.), The emergent lexicon . New York, NY: Academic Press.
- Camaioni, L., & Longobardi, E. (1995). Nature and stability of individual differences in early lexical development of Italian-speaking children. First Language, 15((44/2)), 203-218 IS: 01427237.
- Carlson-Luden, V. (1979). Causal understanding in the 10-month-old. Unpublished Ph. D., University of Colorado at Boulder.
- Choi, S., & Gopnik, A. (1993). Nouns are not always learned before verbs: An early verb spurt in Korean. In E. V. Clark (Ed.), The Proceedings of the 25th annual Child Language Research Forum . New York, NY: Cambridge University Press.
- Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. Journal of Child Language, 22(3), 497-529.
- Clark, E. V. (1993). The lexicon in acquisition. New York, NY: Cambridge University Press.
- Dore, J. (1975). Holophrases, speech acts and language universals. Journal of Child Language, 2, 21-40.
- Dore, J. (1985). Holophrases revisited: Their 'logical development from dialogue. In M. D. Barrett (Ed.), Children's single-word speech . New York, NY: Wiley.
- Dromi, E. (1987). Early lexical development. San Diego: Singular Pub. Group.
- Dromi, E. (1999). Early lexical development. In M. Barrett (Ed.), The development of language . Hove: Psychology Press.

- Dunham, P. J., Dunham, F., & Curwin, A. (1993). Joint-attentional states and lexical acquisition at 18 months. Developmental Psychology, 29(5), 827-31.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. Monogr Soc Res Child Dev, 59(5)(Serial no. 242), 1-173.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. K. II (Ed.), Language development (Vol. 2,). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. Journal of Child Language, 21(1), 125-155.
- Gopnik, A., & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello & W. E. Merriman (Eds.), Beyond Names for Things: Young Children's Acquisition of Verbs . Mahwah, NJ: Lawrence Erlbaum Associates.
- Griffiths, P. (1986). Early vocabulary. In P. Fletcher & M. Garman (Eds.), Language acquisition: Studies in first language development (pp. 279-306). Cambridge: Cambridge UP.
- Griffiths, P. D. (1985). The communicative functions of children's single-word speech. In M. D. Barrett (Ed.), Children's single-word speech . New York, NY: Wiley.
- Halliday, M. A. (1975). Learning how to mean: Explorations in the development of language. New York, NY: Elsevier.
- Huttenlocher, J., & Smiley, P. (1987). Early word meanings: The case of object names. Cognitive Psychology, 19(1), 63-89.
- Kobayashi, H. (1998). How 2-year-old children learn novel part names of unfamiliar objects. Cognition, 68(2), B41-B51.
- Kuczaj, S., A. (1999). The world of words: Thoughts on the development of a lexicon. In M. Barrett (Ed.), The development of language . Hove: Psychology Press.
- Landau, B., Smith, L., & Jones, S. (1998). Object perception and object naming in early development. Trends in Cognitive Science, 2(1), 19-24.
- Le Normand, M. T. (1986). A developmental exploration of language used to accompany symbolic play in young, normal children (2-4 years old). Child, Care, Health and Development, 12, 121-134.
- Locke, J. L. (1996). Why do infants begin to talk? Language as an unintended consequence. Journal of Child Language, 23, 251-268.
- Macnamara, J. (1972). Cognitive basis of language learning in infants. Psychological Review, 79, 1-13.
- Macnamara, J. (1984). Names for things: A study of human learning. Cambridge, MA: Bradford.
- MacWhinney, B. (1995). The CHILDES project : tools for analyzing talk. (2nd ed.). Hillsdale, N.J.: L. Erlbaum.
- MacWhinney, B., & Snow, C. E. (1985). The Child Language Data Exchange System. Journal of Child Language, 12, 271-296.
- Markman, E. M. (1989). Categorization and naming in children. Cambridge, MA: Bradford Books.
- Markman, E. M. (1994). Constraints on word meaning in early language acquisition. Lingua, 92, 199-227.
- McShane, J. (1979). The development of naming. Linguistics: An Interdisciplinary Journal of the Language Sciences, 17, 879-905.
- McShane, J. (1980). Learning to talk. New York, NY: Cambridge University Press.

- Nelson, K., Hampson, J., & Shaw, L. K. (1993). Nouns in early lexicons: Evidence, explanations, and implications. Journal of Child Language.
- Nelson, K., Rescorla, L. A., Gruendel, J. M., & Benedict, H. (1978). Early lexicons: What do they mean? Child Development, 49, 960-968.
- Ninio, A. (1992). The relation of children's single word utterances to single word utterances in the input. Journal of Child Language, 19, 87-110.
- Ninio, A. (1993). On the fringes of the system: Children's acquisition of syntactically isolated forms at the onset of speech. First Language, 13(39/3), 291-313.
- Ninio, A., Snow, C. E., Pan, B. A., & Rollins, P. R. (1994). Classifying communicative acts in children's interactions. J Commun Disord, 27(2), 157-87.
- Parisse, C., & Le Normand, M. T. (1997). Etude des catégories lexicales chez le jeune enfant à partir de deux ans à l'aide d'un traitement automatique de la morphosyntaxe. Bulletin d'Audiophonologie, XIII(6), 305-328.
- Parisse, C., & Le Normand, M. T. (in press-a). Automatic disambiguation of morphosyntax in spoken language corpora. Behavior Research Methods, Instruments, & Computers.
- Parisse, C., & Le Normand, M. T. (in press-b). How children build their morphosyntax: The case of French. Journal of Child Language.
- Quine, W. V. O. (1960). Word and object. Cambridge, MA: MIT Press.
- Random House Inc. (1997). Random House Webster's Unabridged Dictionary. New-York: Random House Inc.
- Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. Developmental Psychology, 33(1), 12-21.
- Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. Developmental Psychology, 28(3), 406-13.
- Rondal, J. A. (1985). Adult-child interaction and the process of language acquisition. New York, NY: Praeger Press.
- Rondal, J. A., Bachelet, J. F., & Péré, F. (1985). Analyse du langage et des interactions verbales adulte-enfant. Bulletin d'Audiophonologie, 5(6), 507-536.
- Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. Developmental Psychology, Developmental Psychology.
- Tomasello, M. (1992). First verbs: A case study of early grammatical development. Cambridge: Cambridge University Press.
- Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. Journal of Child Language, 19, 311-333.
- Tucker, M., & Hirsh-Pasek, K. (1993). Systems and languages: Implications for acquisition. In L. B. Smith & E. Thelen (Eds.), A dynamic systems approach to development: Applications (pp. 359-384). Cambridge: MA: MIT Press.
- Wells, G. (1981). Learning through interaction: The study of language development. New York, NY: Cambridge University Press.
- Werner, H., & Kaplan, B. (1963). Symbol formation: An organismic developmental approach to language and the expression of thought. New York: John Wiley & Sons.
- Woodward, A., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. Developmental Psychology, 30(4), 553-566.

Table 1: Characteristics of the corpora of English children

Origin of the set	Name for the set	Number of children	Age		MLU in words		Percentage of single-word utterances per child		Number of tokens per child	
			M.	SD.	M.	SD.	M.	SD.	M.	SD.
Snow & Pan	Snow 1;2	52	1;2.5	0;0.10	1.08	0.35	83	25	10	11
Wells	Wells 1;6	32	1;6.1	0;1.12	1.39	0.85	74	15	97	68
Snow & Pan	Snow 1;8	48	1;7.28	0;0.28	1.34	0.75	77	17	84	64
Bates & Bretherton	Bates 1;8	28	1;8	0	1.28	0.73	80	21	29	22
Wells	Wells 1;9	32	1;8.25	0;1.6	1.46	0.80	70	16	103	85
Wells	Wells 2;0	31	2;0.1	0;0.18	1.78	1.08	57	21	179	120
Wells	Wells 2;3	32	2;3.3	0;0.10	2.08	1.58	52	21	252	175
Bates & Bretherton	Bates 2;4	30	2;4	0	2.04	1.44	50	19	94	47
Wells	Wells 2;6	31	2;6.0	0;0.5	2.41	1.73	42	15	303	170

Note: the different sets of children are presented in mean age order

Table 2: Characteristics of the corpora of French children

Name for the set	Number of children	Age	MLU in words		Percentage of single-word utterances per child		Number of tokens per child	
		M.	M.	SD.	M.	SD.	M.	SD.
French 1;6	7	1;6	1.12	0.13	88	13	10	4.3
French 1;9	20	1;9	1.57	0.61	61	22	52	85
French 2;0	27	2;0	1.61	0.36	62	17	139	101
French 2;3	24	2;3	2.03	0.65	46	20	189	112
French 2;6	30	2;6	2.62	0.65	34	15	284	179

Note: the different sets of children are presented in mean age order

Table 3: Lexical characteristics of English children's production from different corpora for all type of utterances

Corpus	Snow 1;2	Wells 1;6	Snow 1;8	Bates 1;8	Wells 1;9	Wells 2;0	Wells 2;3	Bates 2;4	Wells 2;6
MLU	1.08	1.39	1.34	1.28	1.46	1.78	2.08	2.04	2.41

Percentages of tokens produced for nouns-and-verbs (N&V) and non-nouns-and-verbs (non-N&V)

N&V	23	47	51	36	49	51	48	36	48
non-N&V	77	53	49	64	51	49	52	64	52

Percentages of tokens produced per lexical category

nouns	19	36	41	30	40	37	29	17	27
verbs	4	11	10	6	9	14	18	18	21
communicators	55	25	19	36	18	15	14	22	11
adverbs	5	4	10	8	7	7	9	7	8
adjectives	3	7	3	2	8	7	6	4	6
wh-words	1	1	2	2	2	1	2	4	2
pointers	3	6	4	3	4	5	5	7	4
others	10	10	11	13	11	14	17	21	22

Percentages of children producing at least one word from a given lexical category

nouns	50	100	96	93	100	100	100	100	100
verbs	25	91	84	50	88	100	100	100	97
communicators	87	97	96	89	100	100	100	100	100
adverbs	19	72	82	57	91	90	97	93	94
adjectives	19	78	67	25	94	100	97	70	94
wh-words	13	25	35	32	25	42	50	60	74
pointers	13	56	61	39	66	90	88	83	97
others	29	91	86	75	97	100	100	97	100

Table 4: Lexical characteristics of English children's production from different corpora for single-word utterances

Corpus	Snow 1;2	Wells 1;6	Snow 1;8	Bates 1;8	Wells 1;9	Wells 2;0	Wells 2;3	Bates 2;4	Wells 2;6
MLU	1.08	1.39	1.34	1.28	1.46	1.78	2.08	2.04	2.41

Percentages of tokens produced for N&V and non-N&V

N&V	25	49	52	36	53	55	41	18	38
non-N&V	75	51	48	64	47	45	59	82	62

Percentages of tokens produced per lexical category

nouns	23	43	49	33	50	48	35	14	32
verbs	2	6	4	3	3	6	6	4	7
communicators	55	29	26	44	23	19	24	56	25
adverbs	7	5	13	8	9	12	20	13	18
adjectives	3	5	3	1	8	8	7	2	8
wh-words	1	1	1	1	1	1	1	4	3
pointers	1	5	2	3	2	2	3	2	2
others	8	5	3	6	4	4	4	4	5

Percentages of children producing at least one word from a given lexical category

nouns	48	100	96	89	100	100	100	80	100
verbs	12	78	55	21	56	74	78	47	77
co	79	97	96	89	100	100	100	100	100
adverbs	19	66	78	50	75	84	97	83	94
adjectives	13	63	51	11	78	90	84	23	81
wh-words	8	16	12	18	13	26	31	27	48
pointers	6	38	31	21	34	42	47	40	61
others	12	69	33	25	69	71	69	37	81

Table 5: Lexical characteristics of French children's production from different corpora for all type of utterances

Corpus	French 1;6	French 1;9	French 2;0	French 2;3	French 2;6
MLU	1.12	1.57	1.61	2.03	2.62

Percentages of tokens produced for N&V and non-N&V

N&V	43	37	43	45	43
non-N&V	57	63	57	55	57

Percentages of tokens produced per lexical category

nouns	29	19	22	23	18
verbs	14	18	21	22	25
communicators	32	23	16	11	5
adverbs	4	8	9	6	7
adjectives	0	2	3	2	3
relative pronouns	0	3	2	2	2
pointers	20	16	14	13	11
others	1	11	13	20	28

Percentages of children producing at least one word from a given lexical category

nouns	71	76	100	100	100
verbs	29	80	100	100	100
communicators	86	92	100	88	93
adverbs	29	64	96	96	97
adjectives	0	28	78	79	93
relative pronouns	0	44	67	46	53
pointers	86	80	100	100	100
others	14	76	96	100	100

Table 6: Lexical characteristics of French children’s production from different corpora for single-word utterances

Corpus	French 1;6	French 1;9	French 2;0	French 2;3	French 2;6
MLU	1.12	1.57	1.61	2.03	2.62

Percentages of tokens produced for N&V and non-N&V

N&V	49	39	48	43	33
non-N&V	51	61	52	57	67

Percentages of tokens produced per lexical category

nouns	32	23	29	24	15
verbs	17	16	18	20	18
communicators	34	33	24	24	19
adverbs	3	7	10	11	21
adjectives	0	0	2	1	2
relative pronouns	0	2	1	1	1
pointers	14	18	15	19	22
others	0	1	1	1	2

Percentages of children producing at least one word from a given lexical category

nouns	71	68	96	100	100
verbs	29	68	93	92	97
communicators	86	80	96	88	97
adverbs	14	40	85	88	90
adjectives	0	8	33	29	47
relative pronouns	0	12	22	21	17
pointers	57	64	93	96	100
others	0	16	26	17	50

Table 7: Most frequent words in all the corpora for English and for French children

English		French					
All utterances	Single-word utterances	All utterances	Single-word utterances				
co oh	9	co uh	9	adv:neg pas	5	adv:place voilà	5
co no	9	co oh	9	adv:place là	5	co ah	5
pro:dem that	9	co no	9	adv:voici voilà	5	co oh	5
v be&3S	9	co yes	7	co oh	5	n voiture	5
pro:dem there	8	co ah	7	pro:dem ça	5	pro:dem ça	5
co uh	7	pro:dem there	6	v:aux être&3S	4	co dodo	4
pro it	6	v look	5	v:aux avoir&3S	4	adv oui	4
det the	6	n mummy	5	pro c'	4	adv encore	4
co yes	6	n daddy	5	n voiture	4	adv:place là	4
wh:pro what	5	n mum	5	det le	4	adv:neg non	4
det a	5	co yeh	4	det la	4	v regarder	3
co ah	5	co yeah	4	co ah	4	n papa	3
v look	4	co um	4	v:mdl aller	3	co coucou	3
pro you	4	co mm	4	pro:subj il	3	v tenir	2
pro me	4	co huh	4	adv:neg non	3	v attendre	2
n mummy	4	wh:pro what	3	v tenir	2	v:pp partir	2
co yeah	4	pro me	3	v regarder	2	v:pp fermer	2
v want	3	n ma	3	v:pp partir	2	n maman	2
v go	3	n car	3	pro:obj l'	2	n bébé	2
pro I	3	co ow	3	n bébé	2	co hein	2

Table 8: Distribution of the twenty most frequent words

	English		French	
	All utterances	Single-word utterances	All utterances	Single-word utterances
Functional words not used in isolation	4	0	9	0
Nouns and verbs	4	6	5	9
Other type of words	12	14	6	11

Table 9: Characteristics of the corpora of English adults

	All utterances		Single-word utterances	
	M	SD	M	SD
Percentages of tokens				
N&V	42	1.2	27	6.5
non-N&V	58	1.2	73	6.5
nouns	14	1.3	19	7.4
verbs	28	1.1	8	2.9
communicators	7	1.2	45	13.3
adverbs	6	1.4	13	6.7
adjectives	5	0.7	5	2.1
wh-words	2	0.3	4	1.8
pointers	5	2.0	3	1.7
others	33	1.8	2	1.4
Percentages of adults producing at least one word from a given lexical category				
nouns	99	1.1	90	9.4
verbs	99	1.1	70	16.9
communicators	99	1.4	97	3.1
adverbs	99	1.1	77	21.4
adjectives	99	1.5	57	23.4
wh-words	83	9.6	52	14.8
pointers	98	2.5	43	22.8
others	100	0.6	26	9.7

Note: standard deviation is computed across corpora.

Table 10: Characteristics of the corpora of French adults

	All utterances		Single-word utterances	
	M	SD	M	SD
Percentages of tokens				
N&V	34	2	27	14
non-N&V	66	2	73	14
nouns	9	2	18	10
verbs	25	1	9	5
communicators	4	0	17	5
adverbs	11	2	32	11
adjectives	4	1	9	3
relative pronouns	8	2	9	6
pointers	3	1	4	3
others	36	2	1	1
Percentages of adults producing at least one word from a given lexical category				
nouns	100	0	84	21
verbs	100	0	74	27
communicators	98	3	86	19
adverbs	100	0	97	4
adjectives	99	1	80	24
relative pronouns	100	0	72	35
pointers	82	26	51	34
others	100	0	47	33

Note: standard deviation is computed across corpora.

Table 11: Most frequent words in all the corpora for English and for French adults

English				French			
All utterances	occ.	Single-word utterances	occ.	All utterances	occ.	Single-word utterances	occ.
pro you	5	co okay	7	v:aux être&3S	6	adv oui	23
v be&3S	5	co oh	6	pro:rel que	5	adv:neg non	10
det the	3	co no	6	pro:subj tu	4	co hein	7
pro it	3	co yeah	5	det le	4	pro:rel quoi	5
det a	2	v look	4	pro c'	4	v regarder&IMP&2S	4
inf to	2	co yes	3	adv:neg pas	3	adv:voici voilà	3
neg not	2	wh:pro what	3	det la	3	pro:rel pourquoi	3
pro:dem that	1	pro:dem here	3	v:aux avoir&3S	3	co mm	2
wh:pro what	1	co peek+a+boo	2	adv oui	2	co ah	2
co oh	1	co hmm	2	adv:neg ne	2	adv :place là	2
v want	1	pro:dem there	2	det un	2	co oh	2
prep in	1	co huh	2	det:dem ce	2	v tenir&IMP&2S	2
pro:dem there	1	co eh	2	pro:subj on	2	v aller&IMP&2P	1
v:aux do	1	co mm	1	pro:subj il	2	co oho	1
pro:dem what	1	co mmhm	1	conj et	2	co ohoh	1
v:aux can	1	co good	1	det les	2	v attendre&IMP&2S	1
conj:coo and	1	adv all+right	1	pro:refl te	1	co bon	1
co no	1	v see	1	pro:subj je	1	v venir&IMP&2S	1
v put	1	co alright	1	prep de	1	co aha	1
v look	1	co whoops+a+daisy	1	prep dans	1	co ahah	1

Table 12: Correlation between children's and adults' production of syntactic categories

Correlation between English children and English adults

Corpus	Snow 1;2	Wells 1;6	Snow 1;8	Bates 1;8	Wells 1;9	Wells 2;0	Wells 2;3	Bates 2;4	Wells 2;6	Mean	SD
All utterances	-0.08	0.15	0.18	0.08	0.17	0.36	0.61	0.71	0.80	0.33	0.29
Single-word utterances	0.97	0.76	0.70	0.95	0.64	0.57	0.74	0.96	0.79	0.79	0.14

Correlation between French children and French adults

Corpus	1;6	1;9	2;0	2;3	2;6	Mean	SD
All utterances	-0.33	0.05	0.30	0.59	0.86	0.29	0.41
Single-word utterances	0.23	0.24	0.35	0.31	0.49	0.32	0.09

Table 13: List of the twenty most frequent words for English children and rank of the words in the adult lexicon

Corpus	Snow 1;2	Wells 1;6	Snow 1;8	Bates 1;8	Wells 1;9	Wells 2;0	Wells 2;3	Bates 2;4	Wells 2;6
No. of non-N&V	14	13	8	14	11	10	9	16	8
No. of frequent words	6	5	6	3	7	7	11	9	12
1	<i>uh</i> (292)	THERE (14)	<i>no</i> (25)	<i>uh</i> (274)	NO (9)	NO (10)	NO (18)	THAT (3)	I (22)
2	<i>woof+woof</i> (457)	OH (16)	<i>yeah</i> (27)	ball (64)	mummy (174)	mummy (173)	I (30)	<i>yeah</i> (28)	NO (12)
3	OH (8)	<i>ah</i> (158)	ball (26)	<i>uuh</i> (437)	da (1219)	<i>oh</i> (21)	THERE (17)	GO (15)	my (80)
4	BALL (18)	<i>mm</i> (73)	mommy (40)	<i>no</i> (26)	OH (14)	THERE (17)	IT (3)	IN (5)	IT (3)
5	NO (16)	<i>da</i> (1207)	A (3)	<i>yeah</i> (21)	THERE (16)	THAT (7)	YES (13)	THERE (13)	want (26)
6	<i>yeah</i> (26)	look (43)	THAT (5)	<i>ah</i> (210)	YES (18)	IT (2)	THE (2)	THIS (11)	OH (14)
7	kitty (134)	<i>uh</i> (337)	IT (4)	THAT (6)	mum (608)	want (24)	<i>oh</i> (21)	<i>here</i> (25)	YOU (1)
8	car (130)	NO (15)	OH (8)	OH (15)	<i>uh</i> (422)	I (27)	THAT (9)	ONE (8)	THE (2)
9	<i>awoh</i> (274)	THAT (10)	I (30)	<i>(o)k</i> (192)	<i>ah</i> (263)	<i>uh</i> (442)	mummy (163)	<i>uh</i> (155)	THAT (9)
10	<i>uhuh</i> (277)	mum	WHAT (10)	<i>uhoh</i> (95)	daddy (320)	me (46)	IN (11)	A (4)	A (4)
11	duck (95)	mummy (221)	THE (2)	bed (37)	me (35)	daddy (404)	YOU (1)	<i>where</i> (57)	ON (5)
12	THIS (11)	ma	duck (86)	mommy (137)	<i>er</i> (1184)	look (38)	look (50)	<i>huh</i> (48)	<i>yes</i> (24)
13	THAT (5)	daddy (239)	chair (76)	<i>where</i> (24)	car (234)	YES (14)	A (4)	OH (14)	GO (20)
14	baby (59)	do (25)	<i>yes</i> (117)	<i>uhuh</i> (333)	A (6)	A (6)	ON (5)	<i>no</i> (45)	look (66)
15	<i>ah</i> (248)	<i>me</i> (45)	car (113)	THERE (14)	THAT (7)	YOU (1)	want (26)	<i>um</i> (390)	me (50)
16	WHAT (13)	<i>ooh</i> (165)	<i>uh</i> (293)	<i>what's</i> (36)	ma (621)	<i>mm</i> (67)	my (90)	I (23)	AND (7)
17	<i>bow+wow</i> (675)	A (5)	<i>there</i> (28)	<i>um</i> (273)	YOU (1)	my (101)	GO (16)	<i>uhhuh</i> (82)	mummy (87)
18	<i>baba</i> (510)	<i>mmm</i> (232)	box (31)	teacher (335)	IT (2)	<i>gone</i> (78)	<i>mm</i> (33)	<i>right</i> (59)	TO (6)
19	mom (184)	<i>gone</i> (100)	kitty (118)	people (34)	<i>down</i> (45)	<i>more</i> (85)	<i>ah</i> (219)	IT (9)	<i>there</i> (21)
20	<i>wow+wow</i>	<i>down</i> (52)	mama (330)	car (40)	<i>mm</i> (66)	<i>mum</i>	<i>mum</i> (644)	WHAT (7)	IN (15)

Note: Words in italics are neither nouns nor verbs, nor purely functional words (determiners, personal pronouns, modals). Words in small capitals and bold face belong to the set of the 20 most frequent adult words. These two criteria are not exclusive.

Table 14: Average rank of the twenty-five most frequent words produced by English children and by English adults

Corpus	Snow 1;2	Wells 1;6	Snow 1;8	Bates 1;8	Wells 1;9	Wells 2;0	Wells 2;3	Bates 2;4	Wells 2;6
Children	122	113	50	91	104	66	50	44	36
Adults	31	33	31	34	31	32	32	39	33

Note: The rank value is computed using as reference the full lexicon of all English adult corpora. The smallest possible value for average rank is 13, which would occur if the 25 most frequent words correspond exactly to the 25 most frequent words in the reference lexicon – average value of 1, 2, 3, ..., 25.

Table 15: Number of word types produced per category by English children for different corpora (all types of utterances)

Corpus	Snow 1;2	Wells 1;6	Snow 1;8	Bates 1;8	Wells 1;9	Wells 2;0	Wells 2;3	Bates 2;4	Wells 2;6
nouns	1.3	24.8	13.7	3.6	31.6	46.6	55.7	9.1	58.4
verbs	0.2	7.4	5.2	1.4	8.0	20.9	36.5	9.5	49.4
communicators	1.9	14.7	4.4	4.5	10.8	16.0	18.8	7.0	17.5
adverbs	0.2	2.8	2.1	0.9	5.1	8.3	12.6	3.5	15.0
adjectives	0.3	5.3	2.1	0.4	6.3	9.3	11.6	2.5	13.7
wh-words	0.1	0.7	0.6	0.4	0.8	1.5	2.8	1.2	4.3
pointers	0.3	4.1	1.2	0.5	2.4	4.8	6.8	2.2	7.0
others	0.3	6.7	4.8	1.6	9.0	20.1	33.9	11.0	49.5

Table 16: Number of word types produced per category by French children for different corpora (all types of utterances)

Corpus	French 1;6	French 1;9	French 2;0	French 2;3	French 2;6
nouns	2.7	9.9	29.8	39.6	47.9
verbs	1.9	5.6	16.2	24.8	38.6
communicators	2.4	7.2	20.1	17.5	11.7
adverbs	0.4	4.0	10.7	12.5	20.0
adjectives	0.0	1.0	4.5	4.8	8.6
relative pronouns	0.0	0.1	0.6	1.5	2.2
pointers	1.6	3.0	4.5	5.7	8.5
others	0.4	16.2	39.0	65.4	124.2

Table 17: Correlations between English and French speakers and between children and adults

Correlations for all utterances

English Children	English Adults	from -0.08 to 0.80
French Children	French Adults	from -0.33 to 0.79
0.77	0.94	

Correlations for single-word utterances

English Children	English Adults	from 0.97 to 0.79
French Children	French Adults	from 0.23 to 0.49
0.74	0.47	

Footnotes

- 1 That, demonstrative pronoun (look at that), is not considered here as a functional word. That, demonstrative article (that car is red), is one and considered as a different word. That, conjunction (the fact that you are here) and relative pronoun (I like the car that is red), are functional words also because these uses do not appear in isolation.
- 2 It may sound strange to correlate the part of speech distribution of two different languages. This makes sense not only because the same system of categories is used, but, more importantly, because the two languages share many grammatical features – especially in their morphosyntactic structure. There are great differences in the detailed uses of these features – which makes it very difficult to write in a foreign language – but not in the general principles.
- 3 Although the most frequent personal pronouns are the same in English and in French, there are some important differences between the two languages. it can be translated in three ways in French: c', il, elle. Unfortunately, il is also the equivalent of he and elle of she, so that if all of these three words can be found in French, a comparison with English becomes difficult. There are also great differences between the children's use of I and you, which are much more frequent than their French equivalent je and tu. Here again, differences between English and French intervene: je and tu are used only as obligatory personal pronouns, and have none of the other functions that I and especially you may have in English. Different words are used for this in French, moi, toi and te.
- 4 In his book, Quine's character is in fact a linguist, not a translator. However, the title of the chapter is 'Translation and Meaning'.