J. Child Lang. **25** (1998), 493–531. Printed in the United Kingdom

Developmental changes and variability in the early lexicon: a study of French children's naturalistic productions*

DOMINIQUE BASSANO
Laboratoire Cognition et Développement, Université Paris V – CNRS

ISABELLE MAILLOCHON
Laboratoire Cognition et Développement, Université Paris V – CNRS

AND

ELSA EME
Laboratoire Langage et Communication, Université de Poitiers

(*Received* 21 *November* 1996. *Revised* 21 *November* 1997)

ABSTRACT

This paper investigates developmental changes, as well as inter-linguistic and inter-individual variations, in the expansion and composition of French children's early lexicons. Two studies were conducted using children's naturalistic productions: a longitudinal study of one child between 1;2 and 2;6, and a cross-sectional study of two groups (12 children each) aged 1;8 and 2;6. Analyses indicate that lexical productivity (measured in types, tokens, and new words) strongly increased with age, whereas lexical diversity showed almost no developmental progression. Nouns and para-lexical elements (including interjections, fillers or formulas) were predominant until 1;8 and decreased over time, while predicates and grammatical words increased. As compared to English, French development was characterized by less frequent nouns, initially more frequent predicates, and a remarkable expansion of grammatical words. Inter-individual variability in lexical productivity, in lexical diversity, and in the proportions of different categories was more marked at 1;8 than at 2;6. Lexical profiles found at 1;8 suggest the

[*] We are grateful to the children and the families who participated in the study. We thank the staff of the nursery schools and our colleagues of the University of Rouen for their cooperation in organizing data collection. We also acknowledge Madeleine Léveillé for assistance in data processing. Address for correspondence: D. Bassano, Laboratoire Cognition et Développement (URA 2143), 28 rue Serpente, 75006 Paris, France. e-mail: dominique.bassano@psycho.univ_paris5.fr

existence of more diversified organizational patterns than those captured in the referential-expressive distinction.

INTRODUCTION

Learning the lexicon seems to be one of the most elementary aspects of language acquisition. However, elaborating the stock of words that form a language involves a complex set of process: children must acquire the phonological forms of words, their meanings, and also their syntactic categories – whether they are nouns, verbs, prepositions, and so on –, as well as the positional and inflectional properties of each of these classes. In this article, we present analyses from a research project which investigates the emergence and early development of the productive lexicon of French children, and focuses on the question of the acquisition of word classes. The present paper deals with two main issues: (1) the quantitative expansion of the first lexicon, i.e. how word production increase evolves in rate during the first stages of language acquisition; and (2) the composition of the first lexicon, i.e. which classes of words comprise young children's productions and how they change across time.

These issues have been largely documented in a number of longitudinal or cross-sectional studies, but almost all of them are based on English-speaking children. The quantitative expansion of the lexicon has been mainly investigated by measuring the number of words that children spontaneously produce, or are said to produce, at different ages. On the basis of these numbers, average tendencies concerning onset time and rate of growth of children's vocabulary could be established. What emerges is that children usually produce their first words at around the average age of one year. Productive vocabulary increases slowly at first, then there is a sudden upturn in the rate of acquisition during the course of the second year of life, around the average age of 1;6–1;8 (the so-called 'vocabulary burst'). For example, in some of the first well known studies in this area, Nelson (1973; 1975) conducted a longitudinal analysis of 18 children's reported vocabulary, and found that these children had acquired ten words at the average age of 1;3, 50 words at the average age of 1;8, and an average of nearly 200 words at the age of 2;0. The upturn in word production occurring between 1;8 and 2;0 is clearly shown in such data. A vocabulary burst was also found in longitudinal case-studies that gave a very accurate picture of a single child's development (Dromi, 1987). A broader picture of children's word production can be found in the large-scale cross-sectional study recently conducted by Bates, Marchman, Thal, Fenson, Dale, Reznick, Reilly & Hartung (1994), and resumed in Bates, Dale & Thal (1995). On the bases of parental reports obtained using the MacArthur Communicative Development Inventories (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick & Reilly, 1993) in which mothers were asked to select from checklists the words their children

produced, Bates and her colleagues examined the vocabularies of 1789 children, ranging in age from 0;8 to 2;6. They found that children had acquired a mean of ten words by 1;0, a mean of 64 words by 1;4, more than 300 words at 2;0 and more than 500 at 2;6 (Bates *et al.*, 1995). However that may be, and as pointed out by Barrett (1995), measures of word production underestimate the rate of early vocabulary growth. It is known that word comprehension anticipates word production, and studies which have compared early word production and comprehension found a gap of several months in the attainment of a similar level of number of words according to language modality (Benedict, 1979; Bates *et al.*, 1995).

General tendencies have also been found in lexical composition, i.e. in the types of words produced by young children and in the way they change during the course of the second and third years. A number of studies have shown stable discrepancies in the order of appearance and frequency of different classes of words. As has long been noted (cf. Grégoire, 1947; Brown, 1973; Braine, 1976; Slobin, 1985), content words form the basis of the one-word stage and of first word combinations in the speech of a majority of children, while closed-class words and inflectional markers usually appear later. In addition, a common finding is that, within the open-class, nouns – particularly, object class names – are learned before verbs and other content words, and initially form the largest part of children's productive vocabulary (Benedict, 1979; Gentner, 1982). Similarly, Bates *et al.*'s (1994) large-scale study identified three 'waves' in the composition and recomposition of children's productive lexicons between the ages of 0;8 and 2;6. Nouns were at first dominant (occupying a mean of 55% of the lexicon of children with vocabularies between 100 and 200 words), but levelled out at around 200 words. Predicates – i.e. verbs and adjectives – started off slowly but showed a constant linear expansion. Finally, closed-class words, which represented a very small proportion of total vocabulary in the period between 1 and 400 words, accelerated sharply relative to other lexical types after that point. According to the authors, these patterns indicate that 'changes in the composition of the lexicon across this development range reflect a shift in emphasis from reference, to predication, to grammar' (1994:98). These cross-sectional findings were in line with the longitudinal results reported earlier by Bates, Bretherton & Snyder (1988), which also showed changes in children's lexicons from a concentration on referential devices to an emphasis on predicative devices culminating in an increased emphasis on the closed-class.

However, studies devoted to the analyses of general changes in the expansion and composition of the lexicon have also pointed out that there is considerable variation among children in the early stages of lexical development. Individual differences have been observed in rate of development as well as in learning style. Variation in the rate of development is apparent

in the range of the number of words acquired at a given age. For example, the Bates *et al*. (1994) study indicated that, by the age of 2;0, the median number of words had reached 311, but the top 10% of the sample was reported to produce 534 words or more, while the bottom 10% was still producing fewer than 57 words. Moreover, some longitudinal studies have shown that not all children display a vocabulary spurt: some acquire words at a steadier rate throughout their second year (Goldfield & Reznick, 1990; Clark, 1993).

As noted by Bates *et al*. (1994), stylistic variations in development are more problematic for universalist models than rate variations, because they call the order and mechanisms of acquisition into question. In terms of lexical development, a well-known dimension of differences between children is the referential-expressive distinction initially proposed by Nelson (1973;1975). In proposing this distinction, Nelson suggested that some children emphasize the referential functions of language, learning words which refer to common objects (i.e. nouns) while others emphasize the social and instrumental functions of language, using many more routines, frozen formulae and other expressive devices, which result in a more self oriented, varied lexicon. The concentration on nouns at the outset of language learning thus appeared not to be such a universal phenomenon. A related form of variability concerns the way in which children come to word combinations. Differences of this type were first analysed in Bloom, Lightbown & Hood's (1975) study: their nominal-pronominal distinction opposed those children who tended to use more substantive terms, particularly nouns, and those who tended to use more pro-forms in their early multi-word utterances. What this distinction suggests is that there is also individual variation in time of appearance and proportion of function words in the early lexicon. These stylistic distinctions – referential vs. expressive, nominal vs. pronominal – have been widely reanalysed and discussed (Peters, 1977; 1986; Bates & MacWhinney, 1987; Bates, Bretherton & Snyder 1988; Lieven, Pine & Dresner-Barnes, 1992; Pine & Lieven, 1993). In particular, a more general contrast opposing 'analytic' to 'holistic' style was proposed to highlight the fact that individual variation could basically be attributed to differences in language information processing mechanisms. According to the analytic-holistic distinction, some children rely most on strategies for extracting small units, whereas others seem to rely on the storage and retrieval of larger but unanalysed input strings (Bates *et al*., 1988, 1994, 1995). Thus, referential and nominal children would be more analytic, producing strings of uninflected content words, while expressive and pronominal children would be more holistic, producing inflections and function words, most likely embedded in unanalysed units.

In their large-scale cross-sectional study of lexical compositions, Bates *et al*. (1994) conducted a quantitative estimation of individual variability in children of the same developmental vocabulary level. These within-

vocabulary-level group analyses confirmed the fact that there was wide variation in referential style, which was over and above the developmental effects previously documented. This variation was most marked in the early stages of development, particularly in children with vocabularies in the 10–50 word range. Children with this vocabulary level had a median 40% of common nouns in their lexicon, but the range extended from 0% to 75%. In contrast with this finding regarding nouns, there was relatively little variation in the proportion of predicates. There was also variation in the closed-class scores for children with vocabularies under 400 words. This was interpreted to represent variation in the analytic-holistic dimension. The study of stylistic variation, thus confirmed overall trends in lexical composition, but also showed that basic words classes vary in time of appearance and frequency, even among children who learn the same language, i.e. English. At the onset of language, there is a particular variation in the proportion of nouns and their role in the lexicon. However, although studies agree that there is a maximum of variability in referentiality in the earliest stages of lexical development which tends to diminish over time, the reasons for this effect are not clear.

Moreover, the extent to which the general tendencies and individual variability observed in English-learning children could also be found in children speaking other languages remains largely to be evaluated. Some studies using parental reports based on Italian (Camaioni, Caselli, Longobardi & Volterra, 1991) and Spanish (Jackson-Maldonado, Thal, Marchman, Bates & Gutierrez-Clellen, 1993) versions of the C.D.I. have been conducted. For example, a study on early vocabularies in Spanish (Jackson-Maldonado *et al.*, 1993) found that, in general, the trajectories of development were very similar for Spanish- and English-speaking children. Productive lexicons of Spanish-speaking children with vocabularies of around 200 words comprised primarily common nouns, and secondarily predicates and closed class items. However, it can be noted that the proportions of predicates and closed class items reported in the first stages of lexical development were higher for Spanish-speaking children (about 22% and 12% respectively) than for the English-speaking children in the Bates *et al.* (1994) study. On the other hand, smaller use of nouns and greater use of other types of items have been reported in studies of free speech for infants who were learning other languages than English. For instance, results of a comparative study of four languages in children around 1;6 indicate that French- , Swedish- , and, above all, Japanese-speaking children had proportions of nouns smaller than English-speaking children (Boysson-Bardies, 1996). Earlier use of verbs and slower growth of nouns (in comparison with English) are suggested also in naturalistic studies on the acquisition of Korean (Gopnick & Choi, 1990; Choi & Gopnick, 1995), an SOV language in which parents often use sentences consisting of a single inflected verb.

These results suggest that more data derived from systematic analyses of languages other than English and using varied methodologies are now needed to assess the extent to which the overall course of lexical development could be influenced by cross-linguistic or cross-cultural differences.

The present research investigates French children's lexical development during the second and third years of life. More precisely, this paper focuses on the two main issues documented above, i.e. early lexical expansion and early lexical composition. We had the double purpose of charting some general developmental tendencies in these processes, and of assessing the extent of variability, from both the inter-linguistic and the inter-individual point of view. In this work, lexical development is approached through a methodology different from that of studies using checklist vocabulary inventories. Analyses of the lexicon are derived from corpus data, i.e. from children's naturalistic productions systematically collected and transcribed. In contrast with vocabulary checklists which document children's 'theoretical' lexicons, data of this type allow for the investigation of what could be considered as their 'actual' lexicons. Although this method limits the number of children under study, it provides rich and diversified data, including, for example, various indices of word production (types, tokens, type/token ratio), as well as formal and contextual properties of the words produced. Cross-linguistic comparison derived from these data will mainly consist of comparing French results to English results, since the latter are the most thorough in the literature and the contrast between French and English presents various points of interest. French is a Romance language generally considered as a language with rich and complex structure. It has a basic SVO word order, but, unlike in English, the canonical order changes from SVO to SOV in a number of cases, e.g. when the direct object is pronominalized. As regard to lexical potentialities, French has a diversified and widely used range of closed-class items of various types. In particular, noun determiners and pronouns show a great functional and morphological richness. They express a variety of distinct meanings, and most of them mark gender, number, and person information (and even some case information for pronouns), resulting in a large variety of words within these classes. Determiners, for instance, involve a number of different definite, indefinite and partitive articles marked for gender and number (*le*, *la*, *les*, *un*, *une*, *des*, *de*, *du*, *au* …, vs. 'a' and 'the' in English). In addition, the use of determiners and subject person pronouns is obligatory in French (unlike in other Romance languages, such as Spanish or Italian), and certain constructions, such as left or right dislocations, which are very frequent in oral French, involve additional pronouns (*moi je veux des fraises* '⟨tonic subj. pron⟩ I want strawberries', *il la prend la balle* 'he ⟨obj. pron⟩ takes the ball') and result in an increase of pronoun use.

Analyses of early lexical expansion and composition presented below are both conducted on data from two complementary studies: a longitudinal

study of one child between the ages of 1;2 and 2;6, and a cross-sectional study of two groups of children aged 1;8 and 2;6 respectively. The longitudinal study was designed to portray a microgenesis of a child's early lexicon, from the emergence of first words to the consolidation of these words into somewhat organized sentences. We wished to determine accurate developmental changes occurring throughout the whole time period. For instance, concerning lexical expansion, we aimed at documenting and accounting for some key points in the increase of word production, such as the well known vocabulary explosion that usually occurs around the age of 1;6 to 1;8. Concerning lexical composition, the central questions were: to what extent can the general course of lexical development observed for English-learning children, that is the 'shift in emphasis from reference to predication to grammar' (Bates *et al.*, 1994), be also found in a French-speaking child? And, are there other types of linguistic elements in early language that play a crucial role in lexical development? The cross-sectional study, which focused on size and composition of children's lexicons at two contrasting moments in the period previously examined, was conducted with a double purpose. First, it was designed to highlight the most obvious changes that occur in the lexicon between these two contrasting and crucial moments. A confirmation and generalization of the main developmental results obtained in the longitudinal study was expected. But we also intended to investigate the extent to which lexical productivity and composition could be affected by individual variation within each age group. As remarked above, a recurrent finding in the English-speaking children literature is that individual differences in rate and style are important at the onset of lexical acquisition (until about 1;6–1;8), and decrease thereafter (Nelson, 1973; 1975; Bloom *et al.*, 1975; Bates *et al.*, 1994). To match this finding, we would expect to observe a greater variability in lexical productivity and composition at 1;8 than 2;6.

METHODOLOGY

Subjects and data collection

The longitudinal corpus. The child (the same as that used in Bassano & Mendes-Maillochon, 1994) was a girl, Pauline, the youngest of four in a middle-class family living in Rouen. At the time of the study she was attending a nursery school. She was audio- and video-recorded twice a month, at home in everyday situations, such as meals, play, washing, dressing, etc., during non-structured interactive sessions (of about two hours each) with her family. Long and non interrupted parts of each of the recorded sessions were exhaustively transcribed, with indications about situations, contexts and gestures, and stored on computer in the CHAT format (MacWhinney, 1995). Parts selected for transcription from the whole session

were chosen so that a certain variety of situations was respected, and a sufficient and representative amount of productions was provided.

In the Appendix (Table A 1), we give detailed information about Pauline's transcribed corpus, using months as units of analysis (combining the two monthly sessions). For each month's data, we calculated the number of productions, i.e. all the child's verbal emissions, and the number of utterances, i.e. those verbal emissions which are linguistic productions. To qualify as an utterance, a production had to be a prosodic and meaningful unit, including at least one element resembling a French word in shape and meaning. Babbling, vocalizations, and completely incomprehensible strings are part of the child's productions, but are not considered utterances and therefore are not analysed in the study. The utterance/production frequency is an index of the child's linguistic productivity, which progressively increased from 0·70 to 0·99 between the ages of 1;2 and 2;6. In addition, the utterance/minute frequency allows us to measure the changes in utterance productivity across time. This index increased from 3 to 6 between the ages of 1;2 and 1;6. By the age of 2;0 it had reached more than 8. Table A 1 also gives information on the child's mean length of utterance (MLU), which increased from 1·1 to 3·3 between the ages of 1;2 and 2;6.

The cross-sectional corpus. Twenty-four children participated in this study: 12 in the youngest group (mean age: 1;8.7; median age: 1;8.8; range: 1;6.18 to 1;9.8), and 12 in the oldest group (mean age: 2;6.0; median age: 2;6.9; range: 2;5.4 to 2;7.11). Each group was composed of 6 girls and 6 boys. All children belonged to middle-class families living in Rouen, and were attending nursery schools either full-time (*crèche*) or part-time (*halte-garderie*).

Each child was individually video-recorded in a special room at Rouen University, during semi-structured sessions (a series of varied play-situations using puzzles, book reading, cubes, etc. were proposed and a gift was offered at the end) in which the child interacted with her mother and two female experiments. For each child, the whole session, which was always preceded by a visit a few days before, had a total duration of about one hour, and the child was recorded for about 45 min. As in the longitudinal study, exhaustive transcripts of selected parts of the sessions were made and stored on computer (CHAT format). In general, transcripts were not as long as in the longitudinal corpus.

Information on the cross-sectional corpus is given in the Appendix (Tables A 2 and A 3). As for each month's sample in the longitudinal corpus, we calculated for each child's sample the number of productions, the number of utterances, and various indices of linguistic productivity. In the 1;8 age group, the utterance/production frequency ranged from 0·74 to 0·95 with a median of 0·84, and the utterance/minute frequency ranged from 3·8 to 11·9

with a median of about 5·5. Both indices were higher and showed a smaller range of individual variation in the 2;6 age-group: the utterance/production frequency ranged from 0·87 to 0·98 with a median of 0·93, and the utterance/minute frequency ranged from 7·7 to 15·6 with a median of about 11·5. In the 1;8 age group, the MLU ranged from 1·27 to 2·11 with a median of 1·57. In the 2;6 group, it ranged from 1·93 to 3·50 with a median of 2·72.

Coding

Analyses of children's lexicons were conducted on monthly (for the longitudinal study) or individual (for the cross-sectional study) samples, each formed of a constant number of 120 utterances. The 120 utterances were selected from each transcribed session, preserving long and non interrupted discursive sequences (but excluding incomprehensible productions when necessary, as explained above). In all, for the 16 months of the longitudinal study, this approach yielded a total of 1920 utterances to be analysed, and, for the 24 subjects of the cross-sectional study, a total of 2280 utterances.

Analysis of the lexicon was based on a specific coding, consisting of a sequential description of all the successive units of each utterance with indications of their grammatical class (e.g. VERB), of their morphological properties if necessary (e.g. PRESENT, INDICATIVE, THIRD PERSON SINGULAR), and of their lexical content (e.g. *manger* 'to eat'). The utterances were coded independently by the first two authors (initial agreement of about 90 %) and discussed until there was complete agreement.

Analyses

Assessment of lexical quantitative expansion. Lexical quantitative expansion was evaluated by mean of a complex indicator that we refer to as 'lexical productivity'. A child's lexical productivity at a given time is obtained through joint consideration of various indices: the number of types and number of tokens produced in the session, and, for the longitudinal study, the number of new words observed in each new session. As is the custom (cf. Pan, 1994), the different phonological forms that could be produced for a word (e.g. *apin* for *lapin*), as well as the differently inflected forms of a word (e.g. *petit* and *petite*, or *mange* and *mangeais*) were considered as the same word-type. In contrast, homophones with distinct meanings and appearing in distinct word classes (such as the article *la* 'the', object pronoun *la* 'her', and locative adverb *là* 'there'), as well as contrasted forms of grammatical words realizing distinct contents (such as the personal pronouns *je*, *me*, and *moi*, or definite articles *le*, *la*, and *les*) were considered to be different word-types.

Taken together, these indices provide a multidimensional evaluation of the child's lexicon at a particular point in time. Number of tokens is a measure of raw word production, number of types indicates vocabulary size, and

number of new words refers to vocabulary enrichment. These indices also give access to more specific measures of lexical diversity, such as the commonly used type/token ratio, i.e. the frequency of different words as a function of total words produced by the child. In addition to this classical index of lexical diversity, we also examine a complementary index, the frequency of new words as a function of types.

Analysis of lexical composition. In order to analyse the composition of the lexicon, we examined which classes of words compose children's language, and which relative place each of them occupies. However, we take into account all the lexical units which form children's naturalistic productions, that is, not only conventional word classes, such as nouns, verbs, and so on, but also those components of children's speech which are not word classes *stricto sensu*, such as sub-lexical and supra-lexical units. Thirteen classes of lexical elements (hereafter referred to as 'words') were thus distinguished in children's productions and were grouped into four main categories: the three conventional categories of nouns, predicates, and grammatical words, and an additional category of what we call 'para-lexical' elements. The first two categories are composed of content words, while the last two consist primarily of elements having a more functional role in the lexicon.

Nouns represent both a class and a category of words. They are elements which are used to designate entities. Nouns include: proper names referring to individuals (*François*, *Popi*); and common nouns referring to classes of animates, namely people and animals (*maman* 'mommy', *bébé* 'baby', *chat* 'cat'), or of inanimates, such as objects, toys, food, body parts, places, abstract entities (*crayon* 'pencil', *pomme* 'apple', *nez* 'nose', *chambre* 'room', *histoire* 'story'). Predicates, which are elements used to tell something about entities, consist of verbs, devoted to designation of actions and states (*manger* 'eat', *être* 'be', *vouloir* 'want'), and adjectives, usually used for the attribution of qualities (*petit* 'small', *mouillé* 'wet'). Grammatical words include a number of function words: adverbs (*là* 'there', *encore* 'again', *puis* 'then'), which were classified in the grammatical word category because they mostly work as function words in these stages of language; noun determiners, such as articles (*le* 'the', *un* 'a'), demonstratives (*ce* 'this'), possessives (*mon* 'my'), numbers (*deux* 'two'), indefinite adjectives (*autre* 'other'), exclamatives (*quel* 'what'), etc; pronouns, which include personal pronouns (*je* '*I*', *moi* 'me', *il* 'he'), demonstratives (*ça* 'that'), relatives (*qui* 'who'), possessives (*le mien* 'mine'), adverbials (*en*), etc; prepositions, i.e. prenominal particles (*à* 'at', *dans* 'in', *sur* 'on'); conjunctions, i.e. particles used to coordinate (*et* 'and', *mais* 'but') or subordinate (*quand* 'when', *si* 'if') phrases; and finally auxiliaries, i.e. the verbs *être* 'be', *avoir* 'have', and *aller* 'go', when they are used in compound verbal forms. The additional para-lexical category is a somewhat heterogeneous group which includes those elements of children's

speech that are not necessarily conventional words. Fillers, i.e. sub-lexical items between babbling and words, which are likely to be precursors of words, are prototypical of this category. We considered as fillers those elements (/a/ in *a chat* 'cat', /ø/ in *eu veux pas* 'don't want'), typically used in a prenominal or a preverbal position, which are phonological approximations of a word and seem to hold the place and function of this word. On the other hand, routines or formulaic expressions, i.e. supra-lexical unanalysed elements (*s'il-te-plaît* 'please', *ça-y-est* 'it's over') are also prototypical of this category. Interjections (*oh!*, *ah!*, *coucou!*), which prototypically convey expression of emotion in an often onomatopoeiac way, can also be considered as part of the para-lexical category. Finally, because the two simple particles used for affirmation (*oui* 'yes') and negation (*non* 'no') are very basic linguistic elements used for interpersonal interactions, we prefer to class these elements in the para-lexical category instead of considering them as adverbs in the grammatical category.

ANALYSES OF EARLY LEXICAL QUANTITATIVE EXPANSION

In this section, we analyse the quantitative characteristics of children's lexical expansion through an evaluation of lexical productivity and diversity. We start with data derived from the longitudinal study in order to give an overall picture of the course of lexical expansion from 1;2 to 2;6. Then we look at the cross-sectional data to have a larger scale characterization of children's lexicons at 1;8 and 2;6, and to evaluate the extent of individual differences for these threshold ages. Since a number of successive analyses have been conducted on both series of data, we present a short discussion of the results within each analysis when necessary.

The longitudinal data: lexical quantitative expansion from 1;2 to 2;6 in Pauline's corpus

Basic indices of lexical productivity. Changes in Pauline's word production are shown in Fig. 1, which presents the quantitative evolution of the three basic indices: number of types, tokens, and new words, calculated for each monthly sample of 120 utterances.

This analysis shows three main results. First, as could be expected, the three indices increased in number across the whole time period, which indicates an overall expansion of the child's lexical productivity. At the age of 1;3, the number of tokens was 136, number of types 38, and number of new words 22. At the age of 2;6, the number of tokens had increased to 392, number of types to 134, and number of new words to 50 (mean numbers by month during the whole period were 209 for tokens, 65 for types, and 26 for new words). Second, the evolution of the three indices was highly correlated: progressions in types and tokens were perfectly correlated ($r = 0{\cdot}97$, $p < 0{\cdot}005$), as were progressions in types and new words ($r = 0{\cdot}97$, $p < 0{\cdot}005$).

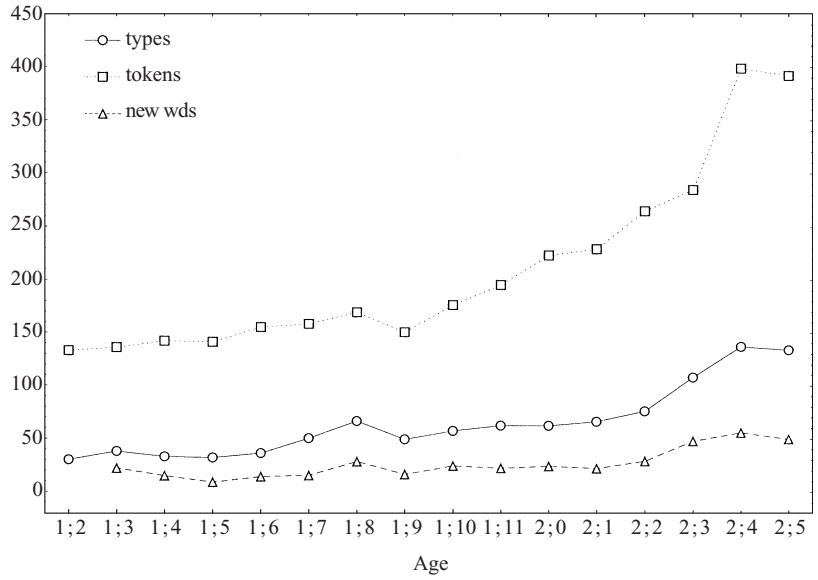

Fig. 1. Basic numbers of types, tokens and new words produced each month by Pauline.

Progressions in tokens and new words were also strongly correlated ($r = 0{\cdot}92$, $p < 0{\cdot}005$). These correlations indicate that the expansion in word production is realized in a remarkably analogous way for the three indices of types, tokens, and new words. It is worth noting that this relation appears not only in the overall evolution, but also in local peaks (e.g. at 1;8) or recessions (e.g. at 1;9). Third, this analysis shows that the lexical expansion was not uniform. Four distinct periods, characterized by a progressively intensified rate of increase, could be found in the evolution of the three indices: a first period of stagnation until the age of 1;6, a second period of slight increase in the rate of progression beginning at 1;6, a third period of marked increase starting from 1;10, and, finally, a sharp increase occurring at around the age of 2;3. This progression can be exemplified in the evolution of type production. The analysis of types showed samples that ranged from 30 to 40 word-types during the first period, from 50 to 60 during the second period, from 60 to 75 during the third period, and finally reached 100 word-types at the age of 2;3.

It is notable that the so-called vocabulary explosion usually observed in children around the age of 1;6 was very attenuated here, whereas more clearcut increases appeared later. We hypothesized that this moderation in increase, which could be specific to Pauline, may be the result of methodological decisions in data collecting and sampling, since monthly samples all consisted of an identical number of utterances, regardless of the child's age.

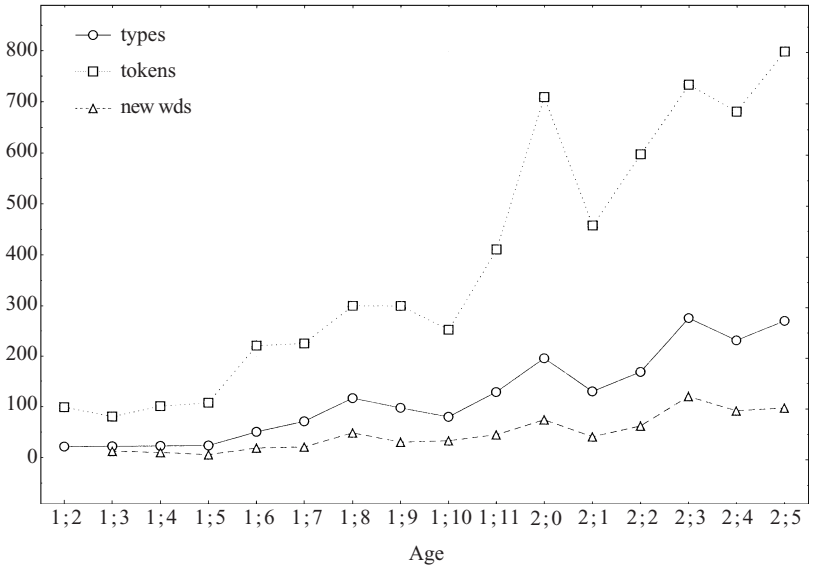

Fig. 2. Weighted numbers of types, tokens and new words produced each month by Pauline.

This method of sampling neutralizes variation across time in the child's language productivity (i.e. the fact that in general a child produces more utterances in month $t+1$ than in month t). A simulation using weighted indices in order to account for these variations was therefore conducted and is presented below.

Weighted indices of lexical productivity. Figure 2 presents the quantitative evolution for the same three indices – types, tokens, and new words – calculated for fictive monthly sessions of equal duration (30 min), in which the number of utterances varies as a function of developmental differences in Pauline's language productivity. Creating fictive sessions of 30 min instead of taking actual real sessions in the corpus was a measure of economy, which, also, allowed comparability with the cross-sectional corpus. Computation of these weighted indices uses the utterance/minute coefficients presented in Table A 1 in order to calculate the theoretical number of utterances produced in 30 min at a given age. For example, the theoretical number of utterances for the fictive session of 30 min at 1;2 is: $2{\cdot}98 \times 30 = 89{\cdot}4$. At this age, the basic number of tokens (for 120 utterances) is 133. The weighted number of tokens will be: $89{\cdot}4 \times 133/120 = 99$. For 30 basic types, the weighted number of types will thus be: $89{\cdot}4 \times 30/120 = 22$.

As can be seen in Fig. 2, this simulation with weighted indices revealed developmental trends which are, overall, similar to those evidenced using the

basic indices. Weighted indices were strongly correlated ($r = 0{\cdot}98$ for tokens and types, $r = 0{\cdot}99$ for types and new words, and $r = 0{\cdot}95$ for tokens and new words, $p < 0{\cdot}005$), and the four distinct periods of word production increase were evidenced. However, there appeared to be some differences between the two kinds of indices. First, overall increases across time were more prominent in weighted indices than in basic indices (for example, the number of tokens increased from 99 at 1;2 to 802 at 2;5, and the number of types from 22 to 274). This is easily explained by the general increase of language productivity with age. More interesting are the particular differences observed. Most striking, the lexical increase at 1;6 was much clearer in weighted indices than in basic ones, thus highlighting the 'vocabulary explosion' phenomenon. In contrast, the increase found at 2;3 using basic indices was quite attenuated in weighted indices. The comparison between patterns obtained with basic vs. weighted indices allows us to see whether the increase in lexical productivity is due to greater language productivity or growth in utterance length. This analysis suggests that the lexical increase at 1;6, corresponding to the classically observed vocabulary burst, is primarily due to an increase in the number of utterances produced. The increase at 1;10 reflects growth in both number and length of utterances, while the increase at 2;3 is mostly the result of an augmentation in the length of utterances.

Indices of lexical diversity. Changes in Pauline's lexical diversity were analysed using two indices: the classical type/token ratio, and the new word-type ratio. This analysis showed two main results. First, in contrast to indices of lexical productivity, indices of lexical diversity did not indicate clear developmental progression across the whole time period. The type-/token ratio, which had a mean monthly value of 0·30, showed a slight increase across time, changing from 0·23 to 0·34. The new word/type ratio was generally higher (mean monthly value of 0·39), but did not increase with age, changing from 0·59 to 0·37. This relative lack of developmental progression for indices of lexical diversity is a logical consequence of the strong correlations observed between progressions of the three basic indices. It shows that, whereas the lexicon expands with age in the absolute number of types, tokens and new words, the process of lexical diversification does not show such a developmental expansion. Second, and also in contrast to indices of lexical productivity, the two indices of lexical diversity were not correlated ($r = 0{\cdot}10$, $p > 0{\cdot}10$), which indicates that differences will be evident through more detailed analyses. These differences are shown in Fig. 3, which presents changes in the type/token and new word/type values across months.

The type/token ratio had a relatively constant value of about 0·23 until 1;6 and showed a sudden increase around 1;6–1;8. But there was no progression after 1;8: from 1;8 to 2;6, there was a relatively constant value of about 0·33. The increase in lexical diversity thus appeared only at the time of the

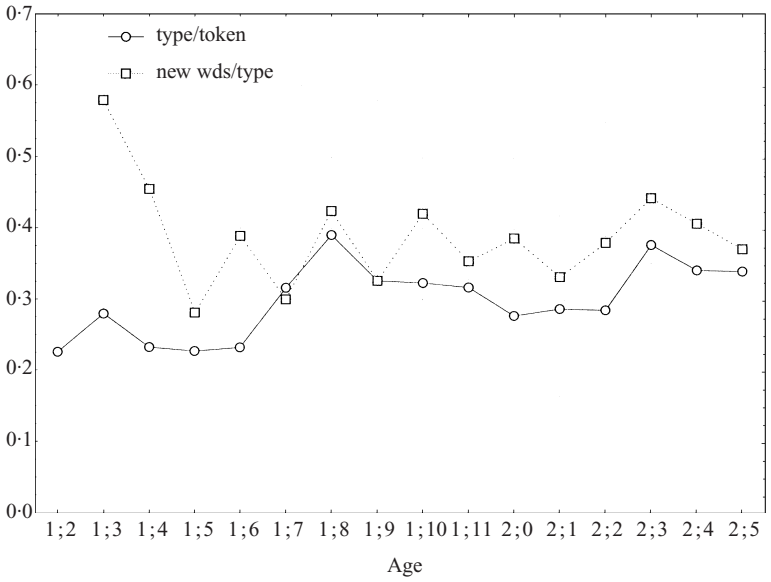

Fig. 3. Evolution of type/token and new word/type ratio across months in Pauline's corpus.

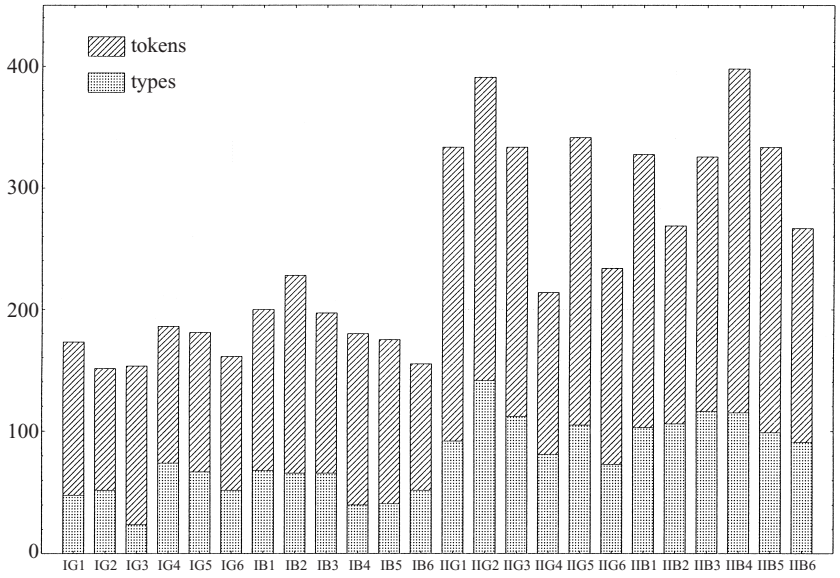

Fig. 4. Basic numbers of types and tokens for children in the 1;8 (IG1 to IB6) to 2;6 (IIG1 to IIB6) groups.

vocabulary explosion. The new word/type ratio gave a quite different profile. The number of new words in proportion to all word-types was particularly high in the first two months, suggesting an initial period of lexical enrichment rather than diversity; after that, from 1;6 to 2;6, it showed an 'up and down' evolution which indicates that inter-month variation in vocabulary enrichment was as important as overall variation between the beginning and the end of the period.

The cross-sectional data : a quantitative picture of children's lexicons at 1 ;8 and 2 ;6

Here, we first examine tendencies found for the various indices of children's lexical productivity and diversity at 1;8 and 2;6, and then we look at within age-group variability.

General tendencies concerning changes in lexical productivity and diversity. As could be expected, analyses of basic indices (cf. Fig. 4 for individual performances) indicate that the level of word production was, on average, considerably higher in the old group than in the young, both for word-types ($t(22) = 7{\cdot}29, p < 0{\cdot}001$), and for word-tokens ($t(22) = 7{\cdot}66$, $p < 0{\cdot}001$). Word production almost doubled between the two target ages: at 1;8, children produced a mean number of 54 word-types and 178 word-tokens, while at 2;6 the mean numbers were 103 word-types and 314 word-tokens. Production of types and production of tokens were positively and significantly correlated at 1;8 ($r = 0{\cdot}62$, $p < 0{\cdot}01$), and at 2;6 ($r = 0{\cdot}81$, $p < 0{\cdot}001$). Analyses of weighted indices (cf. Fig. 5 for individual performances) show even more marked differences between the two age groups in level of word production – mean numbers of types and tokens increased more than threefold. At 1;8, means numbers for a session of 30 min were 89 word-types and 286 word-tokens, and at 2;6 they were 291 word-types and 900 word-tokens. Weighted numbers of types and tokens were still more strongly correlated than basic numbers, at 1;8 ($r = 0{\cdot}96, p < 0{\cdot}001$), and at 2;6 ($r = 0{\cdot}93; p < 0{\cdot}001$). Finally, in contrast to the strong increase in absolute numbers of types and tokens, the slight increase found for the type/token mean ratio between the 1;8 group (0·30) and the 2;6 group (0·33) was not significant ($t(22) = 1{\cdot}23, p > 0{\cdot}10$).

No significant difference in lexical productivity was found between girls and boys collapsing ages ($t(22) < 1$ for types and for tokens). Although boys tended to be slightly more productive than girls, no significant intra-age difference between boys and girls was found. The average basic number of types was 53 for girls and 56 for boys at 1;8, and was 101 for girls and 105 for boys at 2;6. The average basic number of tokens was 168 for girls and 189 for boys at 1;8, and was 308 for girls and 320 for boys at 2;6. There was no effect of sex on lexical diversity: at 1;8 the average type/token ratio was 0·31 for girls and 0·29 for boys, and at 2;6 it was 0·33 for girls and for boys.

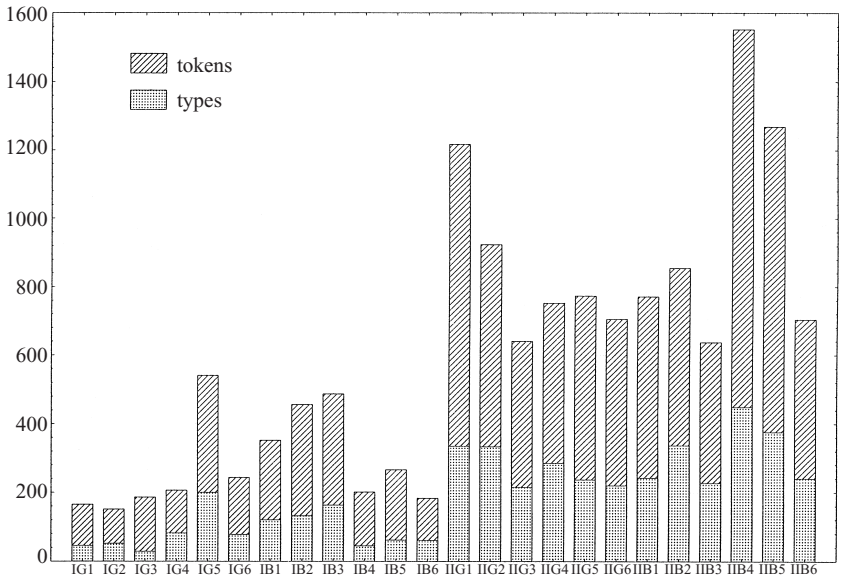

Fig. 5. Weighted numbers of types and tokens for children in the 1;8 (IG1 to IB6) and 2;6 (IIG1 to IIB6) groups.

In summary, a primary finding of these analyses is that, overall, children's lexical productivity strongly increases between the ages of 1;8 and 2;6, whereas lexical diversity shows merely a moderate increase. These general tendencies found in the cross-sectional study confirm the results previously highlighted in the longitudinal study, where it was found that the child's lexical productivity, after an initial period of stagnation, strongly increased between the ages of 1;6–1;8 and 2;6, while her lexical diversity index showed no progression from 1;8 to 2;6. The lack of developmental progression in lexical diversity between 1;8 and 2;6 was also found for the samples in the CHILDES New England corpus (Pan, 1994), in which the mean type/token was lower at 2;6 (0·35) than at 1;8 (0·43).

Intra-age variability in lexical productivity and diversity. An additional question is to determine to what extent there is individual variation in lexical productivity and diversity within each age group. Indicators of variability are given in Table 1 for each distribution, i.e. basic and weighted numbers of types and tokens, and type/token ratio. In addition to mean, minimum, maximum, and standard deviation (S.D.), we calculated the coefficient of variability (CV), i.e. the standard deviation divided by the mean. The coefficient of variability is an index that has been used in a number of studies of variability in various areas of psychology, ranging from the study of animal behaviour to the study of brain-injured patients (e.g. Bekoff, 1977). In the

TABLE 1. *Distribution for basic numbers of types and tokens, weighted numbers of types and tokens, and type/token ratio at 1;8 and at 2;6*

	Basic numbers				Weighted numbers					
	Types		Tokens		Types		Tokens		Type/token	
	1;8	2;6	1;8	2;6	1;8	2;6	1;8	2;6	1;8	2;6
Mean	54	103	178	314	89	291	286	900	0·30	0·33
Min	24	73	151	214	29	215	151	637	0·16	0·28
Max	74	141	228	398	200	449	540	1552	0·40	0·39
S.D.	14·7	17·9	22·6	57·2	53·1	74·3	137·4	291·1	0·07	0·04
CV	0·27	0·17	0·13	0·18	0·60	0·26	0·48	0·32	0·23	0·12

present study it is used to make statistical comparisons of the magnitude of variation across distributions – and, in particular, across age groups – taking differences in the mean into account, and is considered the most relevant index of variability. The 0·30 value is considered here as the point from which the CV indicates an important relative variability.

Let us first consider the variation in basic indices of word-types and word-tokens. As for basic number of word-types, the coefficient of variability was higher for the 1;8 age group (0·27) than for the 2;6 age group (0·17), which indicates a larger relative variability in the younger group. As for basic number of tokens, the coefficient of variability was a little lower for the 1;8 group (0·13) than for the 2;6 group (0·18), but was not particularly high in either group. These analyses indicate that children showed more relative variability at 1;8 than at 2;6 in word-type production, but not in word-token production. In neither age group, however, did the coefficient of variability reach the 0·30 point, either for types or for tokens. Turning now to individual variation in weighted indices, analyses resulted in a larger dispersion, both for types (CV = 0·60 at 1;8 and CV = 0·26 at 2;6) and for tokens (CV = 0·48 at 1;8 and CV = 0·32 at 2;6), particularly in the 1;8 age group. Thus, for weighted indices, both groups showed a great relative variability, which was clearly larger in the 1;8 age group, and was at this age particularly apparent in number of types. Finally, individual variation in lexical diversity can be appreciated through the examination of differences in type/token ratio. These analyses indicate that there was more variability in lexical diversity within the 1;8 age group (CV = 0·23) than within the 2;6 age group (CV = 0·12).

The results regarding individual variation in lexical productivity and diversity can be summarized as follows. First, although raw values for range and standard deviation were higher in the 2;6 group than in the 1;8 group for almost all indices, this is a consequence of the considerably higher production of words at 2;6. Comparative assessment of variability using the coefficient of variability provides a quite different picture. At 1;8, coefficients of variability were higher for basic number of types, weighted number of types and tokens, and type/token ratio. This indicates that children showed larger relative variability at 1;8 than at 2;6 for all indices, except basic number of tokens. The greater variation among the youngest children was most apparent in weighted indices, particularly in weighted number of word-types. If we keep in mind that, in the present study, weighted indices reflect children's linguistic productivity, weighted number of types is the index which most closely resembles the classical measure of vocabulary size (i.e. number of word-types that children reportedly produce). The great variability in weighted types, which we found only in the 1;8 age group, is in line with the recurrent finding in the literature (e.g. Bates *et al.*, 1994) that there is a considerable variability in vocabulary size in very young children

– maximal in the period from 1;6 to 2;0 – and that variability diminishes thereafter. The present study shows, in addition, that variability is more considerable at 1;8 than at 2;6 not only for lexicon size, but also for lexical diversity. However, what also emerges from the present analyses is that, although coefficients of variability were generally higher at 1;8 than at 2;6, they were not particularly high for the basic indices, regardless of the age group, since they did not reach the 0·30 point. If we take into account that individual variation in basic indices is primarily a reflection of children's variability in length of utterances (basic numbers of types and tokens are calculated for samples identical in number of utterances), we can conclude that there was not an enormous variability in length of utterances within either of the age groups. We must keep in mind that, in the analyses of lexical composition which follow, both age groups presented a relative internal homogeneity in basic indices of word production (at least according to our criterion in identifying words), and, consequently, in MLU, which is a classical index of level of language development.

ANALYSES OF EARLY LEXICAL COMPOSITION

As in the quantitative analysis of lexical expansion, the present section on lexical composition starts with analyses of the longitudinal data, and then provides analyses of the cross-sectional data.

The longitudinal data : lexical structure from 1 ;2 to 2 ;6 in Pauline's corpus

Overall distribution of word categories and word classes. In order to give an idea of the structure of the early lexicon, we first examine how each of the main categories of words defined above – nouns, predicates, grammatical words and para-lexical elements – contribute to the constitution of the lexicon. Table 2 gives the overall distributional characteristics for the four categories: overall proportion and monthly mean of types and tokens, monthly mean of new words, and monthly mean values of type/token and new word/type ratios. Overall proportions of types were calculated by cumulating the sums of word-types found in each monthly sample, in the same way as overall proportions of tokens were calculated.

This analysis shows a contrast between nouns and predicates on the one hand, and grammatical and para-lexical items on the other. Nouns and predicates were characterized by higher frequencies in types than in tokens. They had high diversity indices, evident in the type/token as well as new word/type ratios. This general profile for nouns and predicates was reversed for grammatical words and para-lexical items. These items exhibited higher frequencies in tokens than in types, and relatively weak diversity indices. This contrast shows that, as early as the first stages of language acquisition,

TABLE 2. *Distribution of word categories and word classes in Pauline's corpus*

	Types		Tokens		New words (mean)	Type/token (mean)	New/type (mean)
	%	Mean	%	Mean			
Nouns	27	17·4	20	42·5	9·0	0·41	0·52
Predicates	24	15·3	18	36·0	6·9	0·43	0·45
Verbs	19	12·0	15	30·1	5·1	0·40	0·43
Adjectives	5	3·3	3	5·8	1·8	0·57	0·55
Grammaticals	28	18·4	32	68·0	6·5	0·27	0·35
Adverbs	9	6·1	12·5	26·2	2·1	0·23	0·34
Determiners	7	4·4	5	10·9	1·6	0·41	0·36
Pronouns	7	4·5	10	21·6	1·4	0·21	0·31
Prepositions	3·5	2·1	2·5	5·5	0·7	0·38	0·33
Conjunctions	1	1·0	1·5	2·9	0·3	0·34	0·30
Auxiliaries	0·5	0·3	0·3	0·9	0·1	0·33	0·33
Para-lexicals	20	13·4	30	63·0	3·9	0·21	0·29
Fillers	9	5·7	6	12·9	2·2	0·44	0·39
Formulas	2	1·5	2	3·1	0·4	0·48	0·27
Interjections	6	4·2	7	15·1	1·3	0·28	0·31
Yes/No	3	2·0	15	32·0	0·0	0·06	–

content words (nouns and predicates) are highly diversified but used less often, while functional words (grammatical and para-lexical items) are less diversified but frequently used.

Further analyses are aimed at defining the structure of the lexicon more precisely and examining the contribution of the different word classes. Distributional characteristics of the thirteen classes of lexical items present in Pauline's productions are shown in Table 2. Computations of types and tokens show that the different classes of words fell into three groups according to the part they played in the lexicon: we distinguished classes of high frequency, each forming more than 10% of the lexicon; classes of moderate frequency, forming between 5% and 10% of the lexicon; and classes of low frequency, forming less than 5% of the lexicon. Analyses of types indicated that the high frequency group was composed exclusively of nouns and verbs, with nouns being the most prominent. Six classes fell into the group of moderate frequency: adverbs and fillers, pronouns and determiners, interjections, and adjectives. Finally, the low frequency group consisted of the five remaining classes: prepositions, simple 'yes/no' particles, formulaic expressions, conjunctions, and auxiliaries.

Analyses of tokens showed the following distribution. Four classes fell into the high frequency group: nouns, verbs, simple 'yes/no' particles, and adverbs. Four other classes fell into the group of moderate frequency: pronouns, interjections, fillers, and determiners. The remaining five classes were in the low frequency group: adjectives, prepositions, formulaic

expressions, conjunctions and auxiliaries. It is notable that there was a roughly similar hierarchy in frequency between the different classes for both types and tokens, with the sole exception of the simple 'yes/no' particles. This class was markedly different, since these particles were obviously few in types but were very frequently used.

Finally, analyses of type/token mean ratios gave rather unexpected results. Classes having higher lexical diversity, with a type/token ratio that reached more than 0·40, were formulaic expressions, adjectives, fillers, determiners, and (as could be expected) nouns and verbs. All other classes had a more moderate type/token ratio, about 0·30 in general and even lower for the 'yes/no' particles. This suggests that the weak diversity index of the para-lexical category is mostly due to interjections and 'yes/no' particles. In the grammatical category, weak diversity is found in all classes, except determiners, and is particularly evident in adverbs and pronouns. Analyses of new word-type mean ratios showed high frequency of new words (ratio more than 0·40) for adjectives, nouns and verbs.

Distribution of the four main categories of words across time. We now examine how the part played in the lexicon by the four main categories of words changed across time from 1;2 to 2;6 in Pauline's corpus. Figures 6 and 7

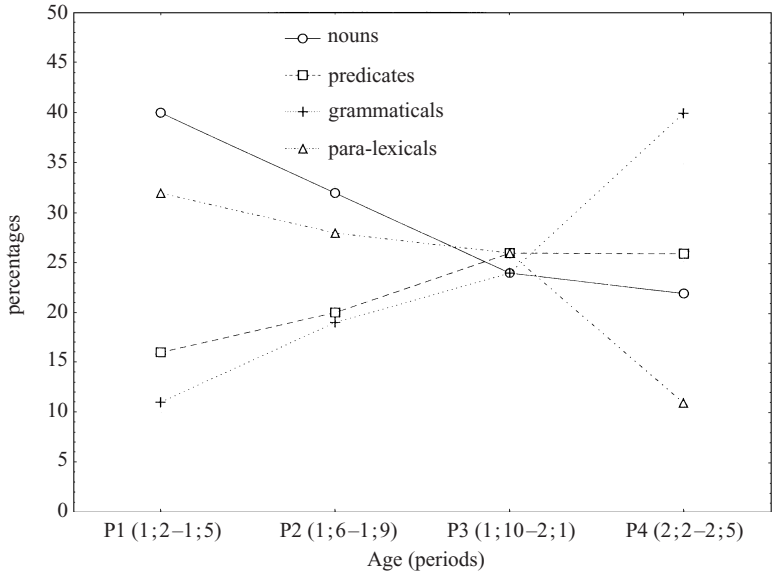

Fig. 6. Proportions of word-types for noun, predicate, grammatical, and para-lexical categories in Pauline's corpus.

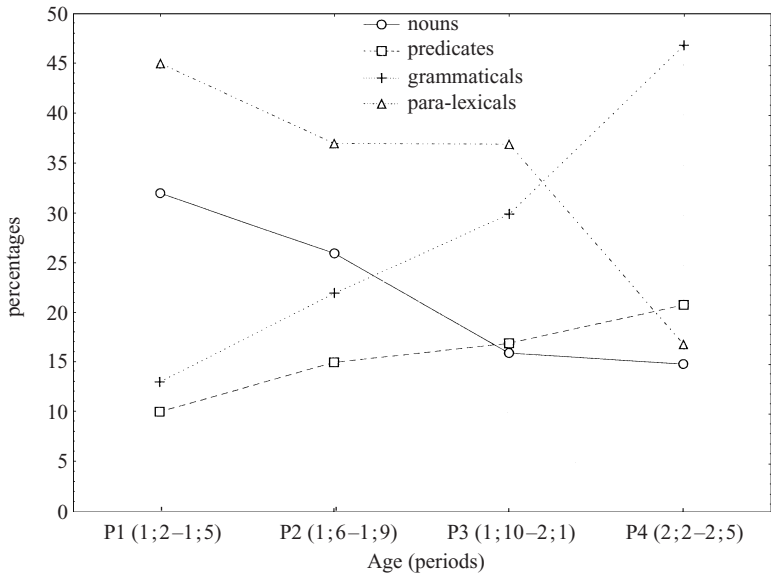

Fig. 7. Proportions of word-tokens for noun, predicate, grammatical, and para-lexical categories in Pauline's corpus.

present the changes in the relative proportions of the four categories across four successive 4-month periods: 1;2 to 1;5 (P1), 1;6 to 1;9 (P2), 1;10 to 2;1 (P3), and 2;2 to 2;5 (P4). These enlarged time segments, roughly corresponding to the four periods previously distinguished in the lexical expansion process, were used because they offer a more clearcut picture of evolutionary tendencies than the monthly segmentation.

Analyses of types (Fig. 6) indicate a striking developmental contrast. Two of the four categories of words, namely nouns and para-lexical items, had initially high frequencies which diminished over time. The other two categories, predicates and grammatical words, exhibited the reverse evolution, with initially low frequencies and subsequently large increases. The four categories of words became proportionately identical (about 25% each) around age 2;0. Before this time, nouns were the dominant category in word-types (forming 40% and 32% of the lexicon in the first and second period respectively), and para-lexical items were the second largest category (forming successively 32% and 28% of the lexicon). Predicates were already present but much less prominent (with proportions of 16% and 20% in the first and second period, respectively), while grammatical words constituted the smallest category of the lexicon before 2;0 (11% in the first period and 19% in the second). After the age of 2;0, grammatical words became the category with the highest frequency, forming 40% of all word-types in P4,

while predicates formed 26 %, nouns 22 %, and para-lexical items 11 % of the lexicon.

Analyses of tokens (Fig. 7) confirm and strengthen the above evolutionary contrast between expanding and diminishing categories: para-lexical items (successively 45 %, 37 %, 37 %, and 17 % of tokens in the four periods) and nouns (32 %, 26 %, 16 %, and 15 %) were initially the most frequent categories and diminished across time, while predicates (10 %, 15 %, 17 %, and 21 %) and grammatical words (13 %, 22 %, 30 %, 47 %) increased across time. When the child was around age 2;0, nouns and predicates were used in equivalent proportions, and when she reached 2;6, nouns, para-lexical items and predicates were used in roughly similar proportions (from 15 % to 20 %), while grammatical words formed a considerable part of all word-tokens. As compared to distribution of types, distribution of tokens showed the following characteristics: para-lexical items, instead of nouns, formed the most frequent category until 2;0, and grammatical words were more frequently used than predicates as early as the first time period and rapidly expanded thereafter.

Results derived from these data indicate that nouns largely dominate over predicates and grammatical words in French lexical acquisition until age 1;8, in proportion of word-types as well as word-tokens. The prominence of nouns progressively diminishes, whereas the proportion of predicates and, above all, the proportion of grammatical words increases across time. Predicates show a rather regular and moderate increase, more marked in types than in tokens. Grammatical words exhibit a strong increase, particularly from the age of about 2;0 on, more marked in tokens than in types. These developments are generally in keeping with results obtained in studies of English lexical acquisition, in particular those of Bates *et al*. (1988; 1994), who found an overall shift in emphasis, over time, from referring elements (nouns), to predictive elements, and, finally, to grammatical elements. The present data also confirm the classic finding that nouns are acquired earlier than verbs (e.g. Gentner, 1982; Maratsos, 1991; Nelson, Hampson & Kessler-Shaw, 1993) – insofar as difference in frequency is a good measure for determining the level of acquisition. However, it is worth noting that verbs are found as early as the first time period (P1) in our data.

Some other more specific results are also derived from the present study of early French lexical composition. First, there is evidence of a preponderance of para-lexical items in Pauline's lexicon before age 2;0. Taken together, these elements were only slightly less frequent than nouns when measured for types, and clearly more frequent than nouns when measured for tokens. Analyses of word class distribution across time indicate that the frequency of types of para-lexical items is primarily due to the prominence of two classes, fillers and interjections. These classes formed, respectively, 11 % and 12 % of types before 1;6. The initial frequency of tokens of para-

lexical items arises, to a certain extent, from the use of interjections and fillers, but is, above all, the result of the frequent use of the 'yes/no' particles, which were largely utilized by Pauline before 1;6 and declined thereafter, particularly in the last months.

The second phenomenon highlighted by the present study concerns grammatical (or closed-class) words. Different development patterns are found for the grammatical category in Bates *et al.*'s (1994) English data and in the present French data, although this category is characterized in both studies by a late expansion when compared to that of nouns and predicates. Bates *et al.* found that closed-class scores were very low (occupying, on average, about 6% of total vocabulary) until children's total vocabulary passed the 400-word point (after the average age of two). After 400 words, the grammatical category started to expand, but it never reached more than 15% of total vocabulary. In contrast, in Pauline's corpus, grammatical words occupy a much more dominant part of the lexicon, forming 10% of the total of word-types as early as 1;2–1;3, about 20% at 1;6, more than 25% at 2;0, and 45% at 2;6. As could be expected, the increase in the frequency of grammatical words is even more noticeable when tokens are considered. Thus, in the present study, grammatical words showed both an early onset and a remarkable expansion: this expansion started as early as around 1;6–1;8, and accelerated strongly after 2;0, in what could be called a 'grammatical explosion' of the child's lexicon. This grammatical expansion, which far surpasses that in the English data, probably reflects a wealth of grammaticality specific to French. Developmental analyses indicate that the child's grammatical category before 1;6 was almost exclusively composed of adverbs (such as *là* 'there', *encore* 'again', *pas* 'not'), and its subsequent increase and explosion was mostly due to the appearance and multiplication of two classes of words, pronouns and determiners. This suggests that the grammatical explosion is marked by a shift in emphasis from adverbs to pronouns and determiners.

The cross-sectional data: a qualitative picture of children's lexicons at 1;8 and 2;6

We now examine, successively, general tendencies and intra-age variability in the composition of the lexicon in the two age-groups. Finally, we examine whether children present distinct 'lexical profiles', defined by the relationships between the four categories of words.

General tendencies in the distribution of word categories. How are the four main categories of words (nouns, predicates, grammatical words, and para-lexical items) represented in the two age-groups? Let us look, first, at the distribution of word categories in absolute numbers of types and tokens. A 3-way ANOVA of word-types (4) word categories $\times$(2) ages $\times$(2) sexes

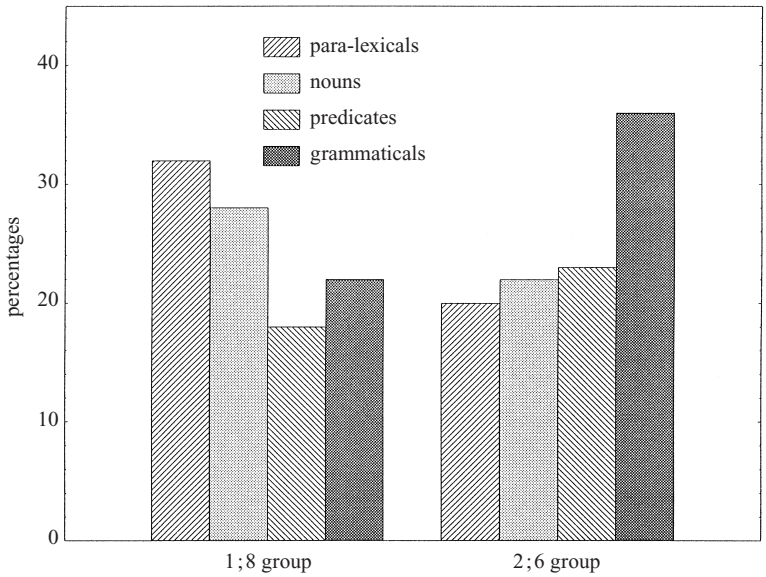

Fig. 8. Mean proportions of word-types for noun, predicate, grammatical, and para-lexical categories in the 1;8 and 2;6 groups.

yielded significant main effects for word category ($F(3–60) = 11{\cdot}10$, $p < 0{\cdot}001$) and for age ($F(1–20) = 48{\cdot}97$, $p < 0{\cdot}001$), but not for sex ($F(1–20) < 1$), and showed a significant interaction between age and categories ($F(3–60) = 19{\cdot}83$, $p < 0{\cdot}001$). Analyses of word categories within each age-group indicate that, at 1;8, children each produced, on average and for samples of 120 utterances: 15 different nouns, 10 different predicates, 12 different grammatical words, and 17 different para-lexical items. At 2;6, they produced 23 nouns, 23 predicates, 37 grammatical words, and 20 para-lexical items. Increase in word-type production between 1;8 and 2;6 was slight but significant for nouns ($t(20) = 3{\cdot}06$, $p < 0{\cdot}01$), stronger for predicates ($t(20) = 5{\cdot}81$, $p < 0{\cdot}001$) and strongest for grammatical words ($t(20) = 7{\cdot}96$, $p < 0{\cdot}001$). It was not significant for para-lexical items ($t(20) = 1{\cdot}39$).

An ANOVA of word-tokens also yielded significant main effects for word category ($F(3–60) = 27{\cdot}48$, $p < 0{\cdot}001$) and for age ($F(1–20) = 58{\cdot}8$, $p < 0{\cdot}001$), but not for sex ($F(1–20) < 1$). There was also a significant interaction between age and categories ($F(3–60) = 21{\cdot}54$, $p < 0{\cdot}001$). In terms of tokens, children at 1;8 produced a mean of 40 nouns, 28 predicates, 49 grammatical words, and 62 para-lexical items, whereas those at 2;6 produced a mean of 48 nouns, 68 predicates, 137 grammatical words, and 62 para-lexical items. Increase in word-token production between the two ages was significant for predicates ($t(20) = 7{\cdot}42$, $p < 0{\cdot}001$) and for grammatical words ($t(20) = 6{\cdot}57$,

$p < 0{\cdot}001$), but not for nouns ($t(20) = 1{\cdot}17$) or para-lexical items ($t(20) = 0$). In both age-groups, the type/token ratio was the highest for nouns (0·37 and 0·47), slightly weaker for predicates (0·36 and 0·34), and weakest for para-lexical items (0·27 and 0·32) and grammatical words (0·25 and 0·27).

Although these analyses of the distributions of absolute numbers provided interesting results, they tended to neutralize the developmental patterns which are revealed by analyses of the relative part each word category plays in the lexicon. Distributions of the four categories of words, in proportion to the lexicon, at 1;8 and 2;6 are shown in Figs 8 and 9. Analyses of word-types (Fig. 8) indicate that the lexicon of children at 1;8 was mostly composed of para-lexical items (32%) and nouns (28%), whereas grammatical words (22%) and predicates (18%) were less frequent. The lexicon of children at 2;6 was composed of an impressive percentage of grammatical words (36%), 22% nouns, 23% predicates, and 20% para-lexical items. Comparison of relative proportions thus shows that the two most frequent categories at 1;8 – namely nouns and para-lexical items – decreased with age, while predicates and above all, grammatical words increased with age. Turning to analyses of word-tokens (Fig. 9) we see that, among all the words produced by children at 1;8, 35% were para-lexical items, 23% were nouns, 27% were grammatical words, and the remaining 15% were predicates. At 2;6, para-lexical items formed only 20% of children's tokens and nouns only 15%, whereas predicates reached 22%, and grammatical words 43%. Although the relative proportions of grammatical words and para-lexical items were more pronounced in the analyses of tokens as compared to types, developmental patterns were analogous: analyses of tokens confirm the contrast between the decrease of nouns and para-lexical items on the one hand, and the increase of predicates and grammatical words on the other.

To summarize these findings, the general tendencies found in the composition of the lexicons of our two age groups of children were perfectly in keeping with the main results obtained in the previous longitudinal study. The present cross-sectional study confirms that between the ages of 1;8 and 2;6, the relative part played by nouns and para-lexical items in children's lexicons (particularly, interjections and fillers) decreased, while the part played by predicates and grammatical words (particularly, pronouns and determiners) increased. Analyses of words produced by the two groups of children serve as further evidence of the preponderant role of para-lexical items in early language, although at 1;8 these elements are probably already less important than they were for younger children. Analyses also confirm more specific aspects of French lexical development. Nouns were less prominent than in English data, and predicates seemed to play a non-negligible part in the early stages. But, above all, analyses point to the early and remarkable expansion of grammatical words. At 1;8, grammatical words already represented about a quarter of the lexicon, with frequencies even

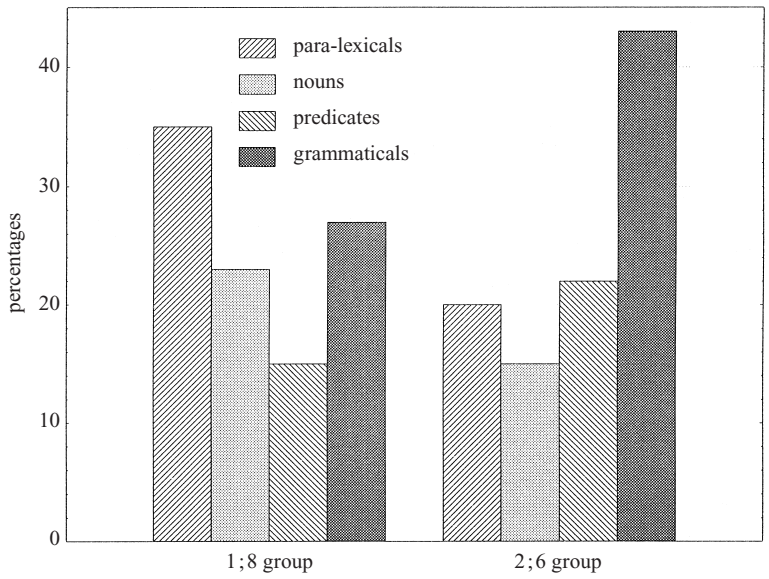

Fig. 9. Mean proportions of word-tokens for noun, predicate, grammatical, and para-lexical categories in the 1;8 and 2;6 groups.

TABLE 3. *Distribution of the four categories of words, for proportions of word-types and word-tokens, at 1;8 and at 2;6*

	Types		Tokens		Types		Token	
	1;8	2;6	1;8	2;6	1;8	2;6	1;8	2;6
	Nouns				Predicates			
Mean (%)	28	22	23	15	18	23	15	22
Min (%)	17	15	11	9	8	17	6	18
Max (%)	40	30	49	21	29	28	29	24
S.D.	0·08	0·05	0·13	0·03	0·07	0·03	0·06	0·02
CV	0·29	0·23	0·57	0·20	0·39	0·13	0·40	0·09
	Grammatical words				Para-lexical items			
Mean (%)	22	36	27	43	32	20	35	20
Min (%)	8	26	5	31	23	13	17	12
Max (%)	31	45	47	55	43	33	50	33
S.D.	0·07	0·06	0·12	0·07	0·06	0·06	0·10	0·06
CV	0·32	0·17	0·44	0·16	0·19	0·30	0·29	0·30

higher than those evident in Period 2 of the child's lexical development in the longitudinal study. At 2;6, they composed about 40% of the lexicon, somewhat less than in Period 4 of the child's lexical development in the longitudinal study.

Intra-age variability in distribution of word categories. To what extent can individual differences be observed in the lexical composition within each age group? To answer this question, we examine variations in the proportion of each category of words. Indicators of variability for the four distributions (proportion of types and tokens in the 1;8 and 2;6 groups) in each word category are given in Table 3.

Analyses of types show that, in the 1;8 group, the proportion of nouns ranged from 17% to 40%, the proportion of predicates from 8% to 29%, the proportion of grammatical words from 8% to 31%, and the proportion of para-lexical items from 23% to 43%. There was thus a similar magnitude of range for all four categories, particularly the first three. The coefficient of variability surpassed the 0·30 point for predicates (0·39) and grammatical words (0·32). It almost reached the 0·30 point for nouns (0·29), but was lower for para-lexical items (0·19). This suggests that, at 1;8, there was considerable relative variability in types for predicates, grammatical words, and nouns. In the 2;6 group, on the contrary, the greatest variability was found in the proportion of para-lexical items, which ranged from a minimum of 13% to a maximum of 33% (CV = 0·30). In the three other categories, variability in the 2;6 group was attenuated when compared to the 1;8 age group: the CV was lower for nouns (0·23), and considerably lower for grammatical words (0·17) and predicates (0·13). Predicates were, at 2;6, the category with the smallest variation in types.

Analyses of tokens showed, in general, greater variability than analyses of types in the 1;8 age-group, but not in the 2;6 age-group. At 1;8, variation in the proportion of nouns was the most marked (CV = 0·57). Variation in tokens was also noticeable for grammatical words (CV = 0·44), for predicates (CV = 0·40), and, to a lesser extent, for para-lexical items (CV = 0·29). In contrast, in the 2;6 age-group, the greatest variability was found in para-lexical items (CV = 0·30). In all three other categories, variability was weaker than in the 1;8 group, showing a CV of 0·20 for nouns, 0·16 for grammatical words, and 0·09 for predicates.

These analyses indicate that, in general, there is more variability in lexical composition at 1;8 than at 2;6. At 2;6, children showed less individual variation than at 1;8 in their proportion of nouns, predicates, and grammatical words. The only exception to this evolutionary pattern was in the production of para-lexical items. The general decrease in variability between 1;8 and 2;6 was apparent in types, and, even more markedly, in tokens. This

was evident in the comparison of absolute ranges as well as in the comparison of the coefficient of variability. This result is in keeping with the idea that stylistic variation is most apparent in the early stages of language development and thereafter decreases in the course of the third year. For example, Bates and colleagues (1994) found that, by the time English-speaking children's vocabularies reach 500 words or more, the variation in the proportion of nouns, predicates, and closed-class words has disappeared. The present study indicates, however, that variability did not completely disappear at 2;6. This was particularly evidenced in the part played by para-lexical items (fillers, interjections), which were still widely used by some children while almost abandoned by others. Variation found at 2;6 in para-lexical items (which are rudimentary linguistic elements) could reflect differences in language maturity and be related to variation in rate of development.

Turning now to the grater variability found in the children at 1;8, we would like to point out the following characteristics. First, variability at 1;8 was typically more evident in proportions of tokens than in proportions of types. This means that children show more differences in the frequency with which they use the different categories of words than in the extension they give to these categories. Second, variability at 1;8 was apparent in at least three of the four categories. To be sure, there was evident variation in the production of nouns – the category upon which a number of studies on early stylistic variability have focused – particularly in the proportion of word-tokens. But early variation was also manifested in the production of grammatical words and predicates. It can be noted that predicates showed relatively little variation in the Bates *et al*. (1994) English data, a difference which could reflect differences in the language being acquired. The extension of variability across word categories suggests that children could follow more diversified patterns than those captured in the referential-expressive distinction when they compensated between the categories of words they preferred to produce.

Looking for lexical profiles. To get a further look at compensation across word categories and see in more detail how children differ in the organization of their whole lexicon, we conducted cluster analyses on the proportions of word categories. Cluster analysis is a technique used to group similar individuals over a predefined set of variables. The method used here for producing clusters is Ward's algorithm, on squared euclidean distances (Ward, 1963). Figure 10 shows the lexical profiles corresponding to each of the clusters found in the analyses conducted on word-types.

This analysis on word-types indicates that the 24 subjects of the study could be classified into four groups (with 73·7% of the variance accounted

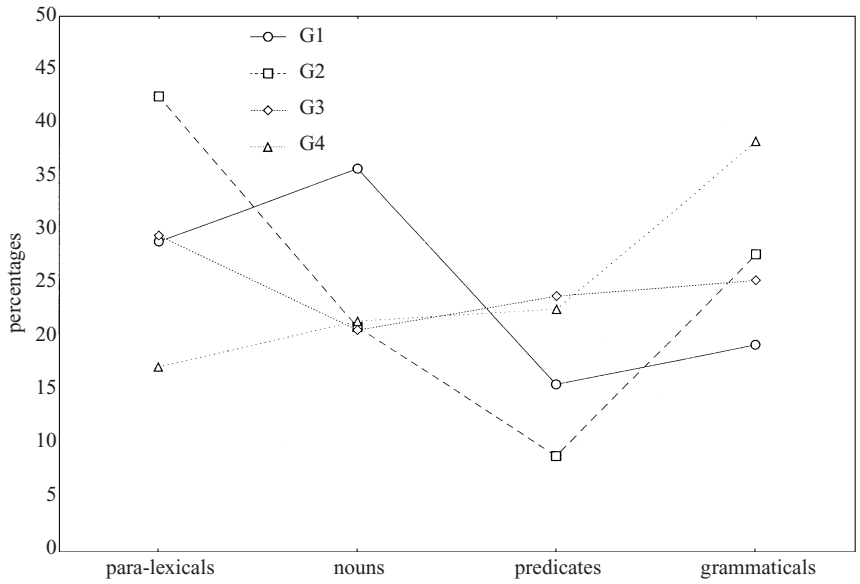

Fig. 10. Lexical profiles: frequencies of word-types categories for each of the four clusters.

for). The profile of the first group, referred to here as the 'referential profile' (G1), was characterized by a particularly high frequency of nouns (a mean of 36%), which was clearly higher than that found in either of the other 3 profiles. The high frequency of nouns is associated with a moderate frequency of para-lexical items and particularly low frequencies of predicates and grammatical words. Among the 6 subjects who have the referential profile, 5 came from the 1;8 group, and 1 from the 2;6 group. In the second profile, referred to here as the 'para-lexical profile' (G2), the low proportion of nouns (22%) was balanced by a particularly high frequency of para-lexical items (43%) and a moderate frequency of grammatical words. This profile was found in 2 subjects, both from the 1;8 group. In the third profile (G3), the low proportion of nouns (21%) was balanced by all other categories, and, in particular, by a surprisingly high proportion of predicates (24%). Among the 7 subjects with this 'diversified predicative profile' 5 came from the 1;8 group and 2 from the 2;6 group. Finally, the last profile (G4), represented by 9 subjects from the 2;6 group, was characterized by a lexical organization of the type PARA-LEXICALS < NOUNS < PREDICATES < GRAMMATICALS, in which the grammatical category was the most prominent (38%). They are said to have a 'grammatical profile'.

This analysis indicates that three distinct profiles must be distinguished to account for the diversity of the lexical organization of children at 1;8, while

the greatest majority of those at 2;6 showed a similar 'grammatical profile', which typifies their age group. It can be noted that the child's lexical structure in the longitudinal study was of the referential type at 1;8 and of the grammatical type at 2;6. These results confirm that qualitative stylistic variability is a real phenomenon at 1;8. Although all of the children at 1;8 produced all four categories of words, they differed in the category they preferred to produce. Although the referential profile, with noun dominance, was frequent (42% of the group), a number of children had fewer than a majority of nouns. However, the non-referential children varied in the way they compensated for their lack of referentiality: some of them (16% of the group) compensated with a majority of para-lexical elements, and most of them (42% of the group) compensated with a majority of predicates associated with para-lexical elements and grammatical words. This suggests that the diversity of lexical patterns found for French children at 1;8 is more complex than that evidenced in the two-part referential-expressive distinction.

GENERAL DISCUSSION AND CONCLUSION

We have presented analyses aimed at investigating the quantitative expansion and composition of the productive lexicon of French children during the early stages of language acquisition. The central purpose of this investigation, conducted using longitudinal and cross-sectional naturalistic data, was to characterize the main developmental changes which occur in these processes between the ages of 1;2 and 2;6, and a further goal was to explore the role of interlinguistic and interindividual variations.

Developmental changes in the quantitative expansion of children's lexicons were evaluated using a complex indicator that we called 'lexical productivity', which refers to changes in word production across constant samples derived from successive video-taped sessions. Analyses of changes in lexical productivity can give a rather complete picture of quantitative lexical development, since they provide information on word-type, word-token, and new-word production, and the derived index of lexical diversity. Moreover, in addition to basic indices obtained using language samples equal in number of utterances, we calculated 'weighted indices', which can give an idea of word production across sessions which are equal in duration and vary as a function of children's rate and level of productivity. One question is to determine how lexical productivity can be related to the classical indices used in studies of early language development, particularly the vocabulary size index. In studies based on parental diaries or reports, vocabulary size refers to the number of distinct words that a child is reportedly able to produce at a given age. In our studies, the evolution of weighted indices – in particular, the weighted number of types – is probably the most in keeping with this

measure of vocabulary growth, whereas the evolution of basic indices – in particular, the basic number of tokens – corresponds to changes in MLU.

On the bases of these criteria, our two studies agree in showing that, between the ages of 1;8 and 2;6, children's lexical productivity strongly increases, while lexical diversity shows almost no developmental progression. In accordance with these results, strong correlations between word-type and word-token production were found in both studies. In addition, the longitudinal study suggests that the period before 1;8 is a period of relative stagnation in lexical productivity and diversity, except for the months just preceding 1;8, when both indicators show a sudden increase corresponding to the so-called 'vocabulary explosion'. It also suggests that the lexical expansion occurring during the period between 1;8 and 2;6 is not regular, showing, in particular, a strong acceleration some time after the beginning of the third year (around 2;3 for the child under study). As noted by Goldfield and Reznick (1996), there is evidence from various areas of language-learning that non-linear change may follow an initial period of linear growth. The vocabulary spurt is a good example of sudden increase. Goldfield and Reznick (1990; 1996) have suggested that it marks a reorganization in children's discovery of language: a rapidly accelerating lexicon indicates that, after a period where words may be learned item by item, children find out that 'all things can be named', thereby discovering something about language itself. The late acceleration that we found in Pauline's lexical productivity could also be a sign of qualitative and reorganizational change, marked by systematization in lengthening utterances and probably related to the 'grammatical explosion' found to occur at this time in analyses of lexical composition. All these findings (and others) could be used to support the view that linear and non-linear trends are found in the developmental process of language acquisition, and that quantitative upturns go hand in hand with qualitative changes and deep reorganizations, a view shared by connectionist models of language learning (e.g. Plunkett & Marchman, 1993).

Regarding developmental changes in the composition of children's lexicons, our analyses show the existence of a variety of lexical units (13 classes) that can be grouped into four main categories: nouns, predicates, grammatical words, and para-lexical items. These categories do not contribute to the constitution of the lexicon in the same manner. Nouns and predicates, i.e. content words, are classes of high density, that strongly contribute to the diversity and enrichment of the lexicon but are less frequent in use, whereas grammatical and para-lexical elements (i.e. functional words) are less diversified classes used with great frequency. The two studies agree in showing a striking developmental contrast between the evolution of nouns and para-lexical elements on the one hand, and the evolution of predicates and grammatical words on the other. While proportions of nouns and para-lexical elements dominate children's lexicons in the first stages and decrease

over time, proportions of predicates and grammatical words increase over time. With respect to the canonical classes of words, these results are generally in keeping with Bates *et al.*'s (1994; 1995) analyses of English-speaking children, which indicate a shift in emphasis, over time, from nouns, to predicates, to grammatical words.

However, in addition to analyses of canonical classes of words, data from the present studies on children's naturalistic productions evidence the important part played by para-lexical elements. In our studies, para-lexical items almost equal nouns in number of types and surpass them in number of tokens in the initial periods of lexical development. The preponderance of the para-lexical elements at the emerging stages of language development, and their decline thereafter, suggests that elements of this kind mark the transition between pre-language and language. In particular, interjections (serving to express and conventionalize emotion), simple particles of affirmation and negation (constituting the first steps in linguistic interaction with others), and fillers (typically serving to prime grammaticalization), are the privileged means for the child to go from pre-linguistic to linguistic systems. The patterns of acquisition observed in this study suggest the following idea regarding lexical (and language) development: children enter language through two pathways simultaneously – the use of content words (employing nouns), and the use of functional words (employing para-lexical elements). These two basic pathways are then reorganized as the children elaborate and enrich them by means of more subtle, appropriate, and diversified linguistic devices – predicates for the content field and grammatical words for the function field.

As for the role of crosslinguistic differences with respect to general developmental tendencies, i.e. interlinguistic variability, questions were raised through the more specific results evidenced in both our longitudinal and cross-sectional studies. Comparison of our French data with the developmental patterns found in Bates *et al.*'s (1994) study on English-speaking children suggests the following differences. Whereas nouns seem to be somewhat less frequent, the part played by predicates, and particularly by verbs, seems to be greater in early French than in early English (until the age of about 2;0). As noted above, similar phenomena have been found in the acquisition of other languages (Choi & Gopnick, 1995). These variations could be related to structural and cultural differences, among which the role of linguistic input seems crucial. For instance, a proportion of about 35% of nouns has been reported in maternal speech for American mothers, vs. 25% for French mothers (Vihman, Kay, Boysson-Bardies, Durand & Sundberg, 1994). However, the most striking difference between our French data and English data concerns the development of grammatical (or closed-class) words. Here we see an early onset and remarkable expansion over time (in types and in tokens), by far surpassing that in the English data. Because the

grammatical explosion is marked by a shift in emphasis from adverbs to pronouns and determiners, a likely explanation of the expansion of grammatical word-types is in the great functional and morphological richness shown by these two classes of words in French. The expansion of grammatical word-tokens can be explained by children's acquisition of certain syntactic constraints, such as the obligatory use of determiners and subject pronouns, and by the development of some very frequent particular constructions, such as left or right dislocations, which involve additional pronouns. The remarkable expansion of the grammatical category that we found in our studies can thus be seen as reflecting the wealth in grammaticality which characterizes French.

Finally, the question of the role of interindividual variability in lexical productivity and in lexical composition was approached in our cross-sectional study. Comparative analyses of the two age groups, conducted using coefficients of variability, indicate that, overall, individual variation was more marked at 1;8 then at 2;6. In particular, children at 1;8 showed strong differences in indices measuring lexicon size (weighted number of types) and lexical diversity, although they were relatively homogeneous in mean length of utterance. Children at 1;8 also exhibited more variation in lexical composition than at 2;6. They differed more from each other in relative proportions of nouns, predicates, and grammatical words than the older children. The older children still show considerable individual variation, however, in their proportional use of para-lexical items, particularly of fillers. The existence of a considerable early stylistic variability was also apparent in the lexical profiles obtained from cluster analyses. Three distinct profiles were found in the lexical organization of the children aged 1;8: a 'referential profile' where nouns dominated, a 'para-lexical profile' where elementary linguistic elements (such as interjections and fillers) dominated, and a 'diversified predicative profile' characterized by a greater use of predicates. At 2;6, this diversity had disappeared, and a typical 'grammatical profile' was found, reflecting the strong grammaticalization of language which has already appeared at this age in French acquisition. These results regarding individual variation at 1;8 and 2;6 confirm the idea that variability in lexicon is generally more apparent in the early stages of language acquisition and diminishes over the course of the third year (cf. Nelson, 1973; 1975; Bates *et al.*, 1994). However, they also suggest that lexical compensations in young children's language could be more complex than those captured in the well-known 'referential-expressive' distinction. This could indicate, among other things, that crosslinguistic differences influence not only general developmental tendencies, but also individual variation.

These analyses could explain, in part, why variability in lexical composition diminishes overall between the ages of 1;8 and 2;6. We would argue that the disappearance of individual variation reflects the children's integration of

certain strong constraints of the linguistic system. Grammaticalization is an example of such a linguistic constraint: when a certain level in language acquisition and lexical development is reached, all children have to use closed-class words to achieve sentence structuration. This interpretation is in line with differential models of cognitive development, which propose that individual differences reflect preferences in processes evoked to fulfil a given function, and that the increase of situational constraints or pressures limits variability (Lautrey, 1995). Such an explanation could also account for a part of the 'late' variability found in language development, e.g. the fact that there is more variation in MLU at 2;6 than at 1;8. Children at 1;8 have just begun to use word-combinations and cannot differ strongly from each other in MLU. In contrast, at 2;6, the utterance has become long enough to allow for the appearance of individual variation, but is still not long enough for variation to be limited. Along the same lines, in a study examining French children's word class acquisition (Bassano, 1996;1998), we find that, although variability in proportion of nouns has almost disappeared at 2;6, children at this age strongly differ in the extent to which they have integrated the grammatical constraint of the noun-class in French, i.e. the obligatory use of determiner before noun. This late variability can be related to the grammatical explosion occurring around this time. These phenomena of late variability thus suggest that the moments at which maximal variability is observed in language acquisition depend on the nature and course of the developmental process under consideration, and occur during the first stages of this process, although not at the very beginning.

To conclude, this paper has presented an overall picture of developmental tendencies and variability in the expansion and composition of French children's early lexicon. Further investigation, now in progress, is needed to study the characterization and evolution of the different parts of speech that constitute children's language.

REFERENCES

Barrett, M. (1995). Early lexical development. In P. Fletcher & B. MacWhinney (eds), *The handbook of child language*. Blackwell.

Bassano, D. (1996). Early lexical development in French: formal and functional aspect of word classes acquisition. *Abstracts of the VIIth international Congress for the Study of Child Language, Istanbul*.

Bassano, D. (1998). Sémantique et syntaxe dans l'acquisition des classes de mots: l'example des noms et des verbes en français. *Langue Française* **118**, 26–48.

Bassano, D. & Mendes-Maillochon, I. (1994). Early grammatical and prosodic marking of utterance modality in French: a longitudinal case study. *Journal of Child Language* **21**, 649–675.

Bates, E. & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.

Bates, E., Bretherton, I. & Snyder, L. (1988). *From first words to grammar: individual differences and dissociable mechanisms*. New York: Cambridge University Press.

Bates, E., Dale, P. & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (eds), *The handbook of child language*. Blackwell.
Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J. & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* **21**, 1, 85–124.
Bekoff, M. (1977). Quantitative studies of three areas of classical ethology: social dominance, behavioral taxonomy and behavioral variability. In B. A. Hazlett (ed.), *Quantitative methods in the study of animal behavior*. New York: Academic.
Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of Child Language* **6**, 183–200.
Bloom, L., Lightbown, L. & Hood, L. (1975). Structure and variation in child language *Monographs of the Society for Research in Child Development* **40**, Serial no. 160.
Boysson-Bardies, B. de (1996). *Comment la parole vient aux enfants*. Paris, Editions Odile Jacob.
Braine, M. D. S. (1976) Children's first word combinations. *Monographs of the Society for Research in Child Development* **41** (1), Serial no. 164.
Brown, R. (1973). *A first language: the early stages*. Cambridge MA: Harvard University Press.
Camaioni, L., Caselli, M. C., Longobardi, E. & Volterra, V. (1991). A parent report instrument for early language assessment. *First Language* **11**, 345–59.
Choi, S. & Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language* **22**, 497–529.
Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: C.U.P.
Dromi, E. (1987). *Early lexical development*. Cambridge: C.U.P.
Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S. & Reilly, J. (1993). *The Mac Arthur Communicative Development Inventories: user's guide and technical manual*. San Diego: Singular Publishing Group.
Gentner, D. (1982). Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In S. Kuczaj (ed.). *Language development*, *Volume 2*. Hillsdale, NJ: Erlbaum.
Goldfield, B. & Reznick, S. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *Journal of Child Language* **17**, 171–83.
Goldfield, B. & Reznick, S. (1996). Measuring the vocabulary spurt: a reply to Mervis & Bertrand. *Journal of Child Language* **23**, 241–46.
Gopnik, A. & Choi, S. (1990). Do linguistic differences lead to cognitive differences? A crosslinguistic study of semantic and cognitive development. *First Language* **10**, 199–215.
Grégoire, A. (1947). *L'apprentissage du langage*. Bibliothèque de la Faculté de Philosophie et Lettres de l'Université de Liège.
Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E. & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language* **20**, 523–49.
Lautrey, J. (1995). Introduction. In J. Lautrey (ed.), *Universel et différentiel en psychologie* Paris: PUF, 1–14.
Lieven, E., Pine, J. M. & Dresner-Barnes, H. (1992). Individual differences in early vocabulary development: redefining the referential-expressive distinction. *Journal of Child Language* **19**, 287–310.
MacWhinney, B. (1995). *The CHILDES project: tools for analysing talk*. Hillsdale, NJ: Erlbaum.
Maratsos, M. (1991). How the acquisition of nouns may be different of that of verbs. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch & M. Studdert-Kennedy (eds), *Biological and behavioral determinants of language development*. Hillsdale. NJ: Erlbaum.
Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* **38** (1–2), serial no. 19.
Nelson, K. (1975). The nominal shift in semantic-syntactic development. *Cognitive Psychology* **7**, 461–79.

Nelson, K., Hampson, J. & Kessler-Shaw, L. (1993). Nouns in early lexicons: evidence, explanations, and implications. *Journal of Child Language* **20**, 61–84.
Pan, B. A. (1994). Basic measures of Child Language. In J. L. Sokolov & C. E. Snow (eds), *Handbook of Research in Language Development Using CHILDES*. Lawrence Erlbaum Associates.
Peters, A. (1977). Language learning strategies: does the whole equal the sum of the parts? *Language* **53**, 560–73.
Peters, A. (1986). Early syntax. In P. Fletcher & M. Garman (eds), *Language acquisition*. (*2ème ed*). Cambridge: C.U.P.
Pine, J. M. & Lieven, E. (1993). Reanalysing rote learned phrases: individual differences in the transition to multi-word speech. *Journal of Child Language* **20**, 551–72.
Plunkett, K. & Marchman, V. A. (1993). From rote learning to system-building: the acquisition of morphology in children and connectionist nets. *Cognition* **48**, 21–69.
Slobin, D. (1985). Crosslinguistic evidence for the language making capacity. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*. Erlbaum.
Vihman, M. M., Kay, E., Boysson-Bardies, B. de, Durand, C. & Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to one-year-old children. *Developmental Psychology* **30** (5), 651–62.
Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**: 236.

APPENDIX

TABLE A1. *Pauline's longitudinal corpus: characteristics of monthly transcribed sessions from 1;2 to 2;6*

Age	Duration of session (min)	No. of productions	No. of utterances	Frequency utter./prod.	Frequency utter./min	MLU
1;2	46	195	137	0·70	2·98	1·12
1;3	56	183	134	0·73	2·39	1·15
1;4	45	148	129	0·87	2·87	1·21
1;5	60	212	186	0·88	3·10	1·20
1;6	29	196	166	0·85	5·72	1·29
1;7	40	306	229	0·75	5·73	1·42
1;8	33	266	235	0·88	7·12	1·46
1;9	32	300	257	0·86	8·03	1·33
1;10	67	428	386	0·90	5·76	1·50
1;11	40	374	339	0·91	8·48	1·83
2;0	30	408	383	0·94	12·77	1·93
2;1	48	413	386	0·93	8·04	2·00
2;2	43	408	390	0·96	9·07	2·27
2;3	41	437	424	0·97	10·34	2·47
2;4	58	410	398	0·97	6·86	3·39
2;5	50	415	409	0·99	8·18	3·36
Mean	45	319	287	0·90	6·71	1·74

TABLE A2. *The cross-sectional corpus : characteristics of each child's transcribed session, in the 1 ;8 age-group (I)* (G = Girl; B = Boy)

Subject (with age)	Duration of session (min)	No. of productions	No. of utterances	Frequency utter./prod.	Frequency utter./min	MLU
I.G1 (1;6.19)	33	154	126	0·82	3·82	1·52
I.G2 (1;8.08)	35	190	140	0·74	4·00	1·27
I.G3 (1;8.30)	35	210	170	0·81	4·86	1·35
I.G4 (1;8.23)	27	126	120	0·95	4·44	1·59
I.G5 (1;8.30)	17	237	203	0·86	11·94	1·59
I.G6 (1;8.09)	30	206	181	0·88	6·03	1·37
I.B1 (1;8.07)	25	229	176	0·77	7·04	1·73
I.B2 (1;7.21)	22	208	176	0·85	8·00	2·11
I.B3 (1;7.28)	18	211	178	0·84	9·89	1·63
I.B4 (1;7.08)	30	154	134	0·87	4·47	1·54
I.B5 (1;9.04)	21	159	128	0·80	6·09	1·46
I.B6 (1;9.08)	34	184	160	0·87	4·71	1·36
Mean	27	189	158	0·84	6·27	1·54

TABLE A3. *The cross-sectional corpus ; characteristics of each child's transcribed session, in the 2 ;6 age-group (II)* (G = Girl; B = Boy)

Subject (with age)	Duration of session (min)	No. of productions	No. of utterances	Frequency utter./prod.	Frequency utter./min	MLU
II.G1 (2;5.10)	14	221	204	0·92	14·57	2·53
II.G2 (2;6.11)	22	216	208	0·96	9·45	3·50
II.G3 (2;6.00)	24	207	183	0·88	7·67	2·83
II.G4 (2;6.07)	16	236	225	0·95	14·06	1·93
II.G5 (2;5.04)	26	241	235	0·97	9·04	2·86
II.G6 (2;5.05)	18	228	217	0·95	12·05	2·07
II.B1 (2;6.15)	20	216	188	0·87	9·40	2·80
II.B2 (2;6.11)	17	244	216	0·88	12·71	2·40
II.B3 (2;5.07)	28	224	219	0·98	7·82	2·63
II.B4 (2;6.17)	15	250	234	0·94	15·60	3·41
II.B5 (2;7.01)	15	250	228	0·91	15·20	2·91
II.B6 (2;7.11)	20	236	211	0·89	10·55	2·34
Mean	20	231	214	0·93	11·51	2·68