Local and global characteristics in the development of morphosyntax by French children*

CHRISTOPHE PARISSE   } *Institut National de la Santé et de la*
MARIE-THÉRÈSE LE NORMAND   } *Recherche Médicale (INSERM), Paris*

ABSTRACT

Statistical analysis of morphosyntax acquisition in French children aged two to four can be conducted at two levels. At the local level, regularities in the immediate lexical context of a word are evaluated by computing the proportion of pairs of consecutive words (e.g., such as) produced by children which are also produced in the same order by adults. At the global level, regularities in collected samples of children's language production are evaluated by computing the correlation between child distribution and adult distribution of open-class lexical categories in language production samples. Data from child language production show that irregularities evolve at the global level and tend to disappear over time. In contrast, local regularities already present at the age of two remain stable during maturation. To account for these findings, it is suggested that: (a) local regularities are a consequence of the reproduction of the adult's morphosyntax, and global irregularities a consequence of the randomness of a child's cognition and understanding of the world; and (b) high-level syntax – as opposed to morphosyntax – appears only later, as a development of local-level rules, and global regularities are a consequence of high-level syntax.

INTRODUCTION

In early language development, it is very important to follow the course

---

of acquisition using linguistic indicators that change with age. For example, knowing that overgeneralization errors occur after a prolonged period of correct use, or that a large number of nouns are produced before the advent of verbs, is significant for theories about rule learning and cognitive development. The first goal of the present study is to show that different aspects of children's grammatical knowledge evolve at varying rates during the course of language acquisition, and to discuss the consequences of this for language acquisition theories. Our second goal is to show that it is necessary for studies of language acquisition to take into account quantitative as well as qualitative data about children's production.

How linguistic features evolve in child language has been investigated in a number of classic longitudinal studies (Bloom 1970, Bowerman 1973, Brown 1973, Nelson 1973). For obvious technical reasons, very few studies have followed large numbers of children over a long period in order to assess statistical data on language production. Seminal work in this field comes from Wells (1985), Fenson, Dale, Reznick, Bates, Thal & Pethick (1994) and Caselli, Casadio & Bates (1999). Fenson *et al.* (1994) have shown that comprehension starts, on average, between 8 and 10 months, production generally starts at 12 to 13 months, with a dramatic acceleration between 16 and 20 months, and grammar's peak growth occurs between 24 and 30 months. Wide variations around the group average are, of course, also observed, even in healthy children where sociocultural index is controlled for. Nevertheless, it is interesting to note that the evolution of these three aspects of language, although distributed over time, follows the same type of developmental curve. At the age of two, several function words have already appeared, but their use will not generalize until the grammatical spurt at 24–30 months. The dynamic pattern of lexicon and syntax acquisition has been presented by Robinson & Mervis (1998), and Kemper, Rice & Chen (1995) have published growth curves for children aged five to ten. In older children, the principles of syntax are often the focus of interest, as is the case with the studies by MacWhinney (1978) and Pinker (1984). In this context, the goal of the current study is not to describe French children's syntactic acquisition in detail, but rather to reveal some of the underlying properties of this acquisition.

Evaluation of statistical regularities

Since the introduction of computerized analysis of child language data, it has become easier and easier to analyse large amounts of data quickly and automatically. This has promoted the statistical analysis of large

databases, which can be performed using various types of information. In this study, the analysis aims to document the development of syntax in children. Using adult production as a reference, we determine the evolving differences between child and adult language production. Arguably, this is no different from checking for children's syntactic errors with reference to standard grammatical usage. However, using real adult data helps to avoid unwanted interpretations of what an ideal syntax should be. Moreover, the same procedure can be used to compare groups of children, which is usually impossible to do with reference to a non-existent prototypic child grammar.

Syntax can be evaluated in many different ways, and we focus here on two complementary analyses, at the local and global levels. The local level is defined as the immediate lexical and syntactic context of a word in a sentence produced by a child or an adult (i.e., the preceding and following words). In languages with more complex morphological properties than French and English, the local level would also include affixes. Examples of local regularities include case-markers, verb structure, the order of function words, and the structure of the noun phrase or of compound words. The local-level analysis is closely related to the distributional regularities used in computational linguistics, especially in the Markovian analysis of language (Charniak 1993) and the study of child language development (Maratsos 1982, Pinker 1987).

The global level is defined as regularities that span more than one utterance (i.e., at the transcript or text level). Examples of global regularities include mean length of utterance (MLU) and frequency of use of determiners or personal pronouns in obligatory contexts. These global-level measures can be taken as a good index of the morphosyntactic development of the child's language. The use of nouns, verbs, adverbs or adjectives can also be derived from global-level analyses. The use of these categories is more related to semantics and high-level syntax than to morphosyntax. Thus, if children's utterances are short – with only one such word per utterance – statistics using these categories would reflect the child's conceptual organization of the world, i.e., whether s/he tends to focus on objects, actions, properties, etc., or the prominent syntactic characteristics of her mother tongue. For utterances with more than one content word, the statistics would reflect the basic organization of a sentence, i.e., the use of main verbs, subjects, objects, etc.

Syntactic analysis of language transcripts

In contrast to traditional approaches focusing on lexical analyses, all

transcripts of linguistic productions were tagged for part of speech, using a set of 25 categories (see Appendix for a complete description of the tag set used). This number of categories corresponds to an average level of syntactic description and allows for a distinction between the generic classes of noun, verb, adjective, adverb, determiner, pronoun, conjunction, preposition, number, communicator. This distinction could be considered as essentially semantic for the first four categories and grammatical for the others. This distribution into 25 categories also allows for subcategories of the classes noun, verb, adverb, determiner, pronoun, preposition, communicator. These subdivisions were made on the basis of either unambiguous lexical differences which correspond to a real difference in use – such as the different classes of adverbs – or differences in the syntactic context of use. For example, generic nouns in French are always preceded by a determiner, whereas this is usually not the case for proper nouns. Another example is that the past particle is usually preceded by the auxiliary 'to have' or 'to be' whereas infinitives are usually preceded by modals. Similarly, the immediate word context would define the difference between auxiliaries and main verbs.

This range of categories thus makes it possible to observe the development of the morphosyntax, a local characteristic of linguistic regularity, and the evolution of content word use, a global characteristic of linguistic regularity. A comparison between child and adult morphosyntactic co-occurrences is done to determine the extent of local regularities. A *morphosyntactic co-occurrence* is defined as two or three words, or two or three syntactic categories, which occur one immediately after the other (and always in the same order) in a sample of language output. In this article, the term *bi-tag* designates two successive syntactic categories, *bi-word* two successive words, and *tri-tag* and *tri-word* three categories and three words, respectively. The term *co-occurrence* designates all these types unless otherwise specified, whereas the term *collocation* is reserved for word co-occurrences of any size. Finally the term *syntactic co-occurrence* will designate co-occurrences of syntactic categories. The comparison between child use and adult use involves computing the percentage of children's bi-tags (also tri-tags, bi-words and tri-words) that can be found in adult corpora. It is also possible to measure the lack of, or the correct use of, pronouns and articles inside and outside obligatory contexts.

The proportion of nouns, verbs and function words used in the children's and adult corpora is computed to determine the extent of global regularities. In this case, the Pearson product-moment correlation

.

between the observations for children and for adults is computed. The higher this correlation is, the more similar are the children's and adults' use of syntactic categories. Correlations can be computed for all syntactic categories, all content word classes, or only for generic categories of content words (nouns, verbs, adjectives, adverbs, communicators, interrogative-connector, object-focus). These generic categories correspond to those produced in isolation by children (Parisse & Le Normand 2000).

METHODOLOGY

Participants

Children aged 2;0 to 4;0 participated in this study. Children were grouped by age because, contrary to what Lieven, Pine & Barnes (1992) suggest for younger children, it is difficult to classify children after age 2;0 by the size of their vocabulary. Spontaneous language samples were collected every three months, creating nine groups of 20 to 30 children. The number of participants, their MLU and the number of utterances produced are presented in Table 1. All participants had a normal linguistic development pattern. Children were excluded from the study if they manifested hearing difficulties or a lack of symbolic play; they were tested using Lowe & Costello's (1976) testing procedure. All children were lingual speakers living in a monolingual environment.

Adult language material

In order to compare the lexical classes of young children with those of adults, it was important to use a corpus as close as possible to ordinary spoken language and in a natural setting such as a conversation with, or in the presence of, a child. The adult reference corpus presented here is taken from the CHILDES database (MacWhinney & Snow 1985) and consists of the whole set of adult data (extracted from conversations with the child Philippe) gathered by Madeleine Léveillé with the participation of Patrick Suppes (Suppes, Smith & Léveillé 1972, Suppes, Léveillé & Smith 1974). This corpus corresponds to 33 one-hour-long tape-recordings of a child at home, and covers a year. The transcripts include the utterances both of the child Philippe and of the adults, namely, the child's mother and the father and field researcher Léveillé. Every sentence was analysed, and the utterances were separated into child and adult utterances. The 'adult corpus' contains 22,669 utterances including 130,053 words, without punctuation and with an MLU of 6.33.

TABLE 1.   *Characteristics of children's corpora*

Age	No. of subjects	MLU		Number of utterances		
		Mean	SD	Total	Mean	SD
2;0	27	1.63	0.37	2184	80	44
2;3	24	2.04	0.65	2180	90	32
2;6	30	2.62	0.65	3179	105	54
2;9	24	3.33	0.79	3324	137	105
3;0	25	3.70	0.80	2883	114	42
3;3	23	3.82	0.49	3473	150	76
3;6	23	4.11	1.28	2907	125	53
3;9	20	4.42	0.48	2212	110	54
4;0	28	5.39	2.19	4052	144	105

Transcription

The child data were collected using the technique of direct observation of a young child's behaviour. Direct spontaneous language was produced during symbolic play, always in the same standard situation, always openly video-recorded, and always by the same observer (Le Normand 1986). The recordings were made in a play situation allowing children to talk about their own actions and real or imaginary events and to interact with a familiar adult partner. The standardized play material involved five toy figures (two adult figurines, two child figurines and one baby), one dog, eleven pieces of furniture (two tables, four chairs, two armchairs and three beds) and five figurative objects (stairs with a mobile door, a garage with a sliding door and a front door bell). The children's speech was then segmented into utterances using the criteria defined by (Rondal & Pérée 1985), which also use the computation of linguistic parameters described in the CLAN corpus processing system (Child Language Analysis, version 2.01, Mac-Whinney 1995).

Procedure

The parts of speech of the corpora used in this study were automatically tagged using POST, a part-of-speech tagger program available with the CHILDES system (see http://childes.psy.cmu.edu, Parisse & Le Normand 1998, 2000). Morphosyntactic parsing was necessary because 2-year-olds can produce ambiguous words. In the

French children's corpus presented below, children aged 2;0 produced 3,765 words, in tokens. Of these words, 534 were used ambiguously by the children – for example, one child used *la* as an article and *la* as an object pronoun – and 2,433 are inherently ambiguous in French, i.e., they have more than one category in the French lexicon so that the children could, potentially, produce them in different categories (Parisse & Le Normand 1998). This aspect and the size of the corpora to be analysed make the use of automatic linguistic analysis techniques mandatory in a study of the similarities and differences between child and adult language.

The principle of POST is to tag texts automatically after as short as possible an automatic training period. It is based on a Markov model of ambiguous bi-tag succession rules. When a choice between several classes immediately follows a previous choice between several other classes, the two-word unit is considered as an ambiguous bi-class, for example, *la porte* (the door): ARTICLE or PRONOUN followed by NOUN or CONJUGATED VERB; or 'her book': OBJECT PRONOUN or POSSESSIVE ADJECTIVE followed by VERB or NOUN. It will also be considered as such if only one of the two words has an ambiguous class status (our book). The ambiguity has to be resolved and one must decide which two classes are correct in the current context. In the previous example, *la porte*, the possibilities are ARTICLE followed by NOUN, and PRONOUN followed by VERB. The context is taken into account in the final decision. With 'her book', the only possibility is POSSESSIVE ADJECTIVE followed by NOUN. Using the succession of ambiguous bi-classes and their rules of resolution, the analyser provides sentence-level resolution. There may be more than one solution, in which case they are sorted out on the basis of the probabilities of grammatical bi-class occurrences. The morphosyntactic analyser is well adapted to solve lexical ambiguity problems, which are very common in French, as in many other languages. A complete description of the tag set used in this article and of class usage by children aged 2;0 and 4;0 and by adults is given in the Appendix.

Analysis

Two different measures were computed for local regularities. First, child language and adult language were compared by computing the percentage of bi-tags, bi-words, tri-tags and tri-words produced by the children that were also produced by the adults. Second, child language and adult language were characterized by computing the ratio of the number of articles produced to the number of nouns produced, and the

ratio of the number of pronouns produced to the number of verbs produced.

Two measures were also extracted for the global regularities. First, the children's language was characterized by computing the proportion of use of nouns, verbs and function words. Second, child language and adult language were compared by computing the correlation between the distribution of syntactic categories produced by the two groups. This correlation is computed separately for all syntactic categories, for content word categories only, and for the seven generic categories.

RESULTS

Local regularities

The percentage of children's co-occurrences that are also adult co-occurrences is very high. For bi-tags, the mean value is 98% (SD 1.6%), for tri-tags 93% (SD 6.8%), for bi-words 64% (SD 10.2%), and for tri-words 35% (SD 14.9%). In spite of the higher percentages for categories over words, the values for collocations remain impressive. Although they originate from different corpora (different situations, different people), at the age of two, 58% in types (63% in tokens) of the children's bi-words are also adults' bi-words, and 42% in types (43% in tokens) of the children's tri-words are also adults' tri-words.

One of the most remarkable results of these statistics is that the percentage of child-adult co-occurrences does not vary much with age. Independently of the children's language development, the proportions remain identical for bi-tags and bi-words, but decrease slightly for tri-tags and tri-words. The effect of age is not statistically significant for bi-tags ($F(8,215) = 0.66$, $p = 0.72$), for tri-tags ($F(8,215) = 1.43$, $p = 0.18$), or for bi-words ($F(8,215) = 1.24$, $p = 0.27$), but is significant for tri-words ($F(8,215) = 2.64$, $p = 0.009$). The coincidence for tri-words at the age of two is 43%, decreasing to 28% at age four. This decrease may be due, in part, to the children's 'taking chances' in creating new grammatical configurations, to overgeneralization errors, or to older children's richer cognitive world, which may be less stereotyped than younger children's. It should be noted that, when we compute correspondence values between Philippe – the only child in the Léveillé corpus – and his mother, the figures go from 77% for bi-words (66% in types) at age 2;1 to 70% (66% in types) at age 3;2, and from 41% for tri-words (36% in types) at age 2;1 to 39% (32% in types) at age 3;2. These figures do not differ greatly from those for the children in the Le Normand database. In addition, the percentage of correspondence between the adults in the Léveillé corpus and the adult
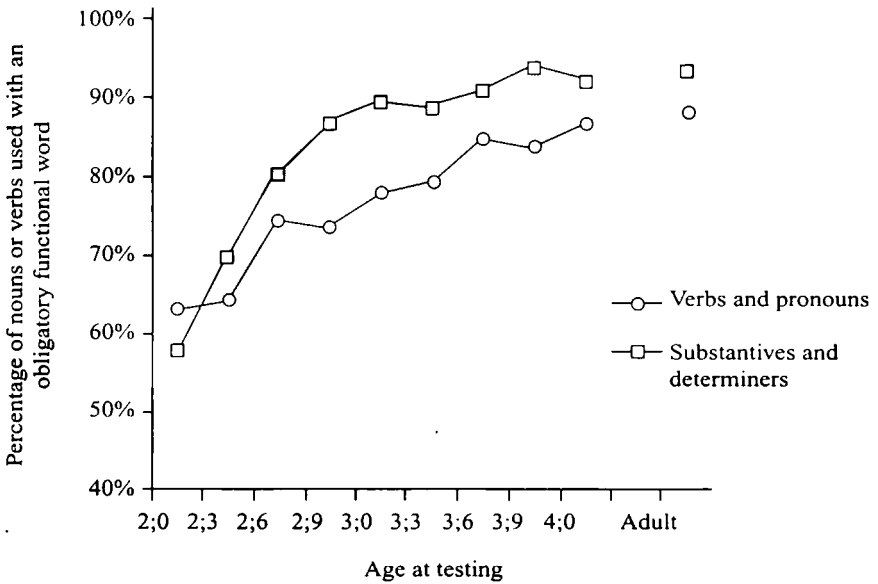
Fig. 1. Ratio of nouns accompanied by a determiner to the total number of nouns, and ratio of verbs accompanied by a pronoun to the total number of verbs (*Note*: this is computed for multi-word utterances only)

in the Rondal corpus (in the CHILDES database) drops to 56% (21% in types) for bi-words and to 25% (9% in types) for tri-words; the greater difference between token and type values is due to the respective size of the two corpora. It thus seems that local characteristics are independent of age and subject. Although there are individual differences between children, they tend to decrease with age. By contrast, the mean percentage of children's collocations that are adult-like does not vary much. The standard deviation for bi-words is reduced by half between the ages of 2;0 and 4;0. This change is still stronger for tri-words. These trends are significant, as revealed by the test for homogeneity of independent variances between the percentage of child bi-words (tokens) in adult bi-words at age 2;0 and the percentage at age 4;0 ($F(26,27) = 3.71$, $p = 0.002$). For types and tri-words, the results are also significant. The results using syntactic
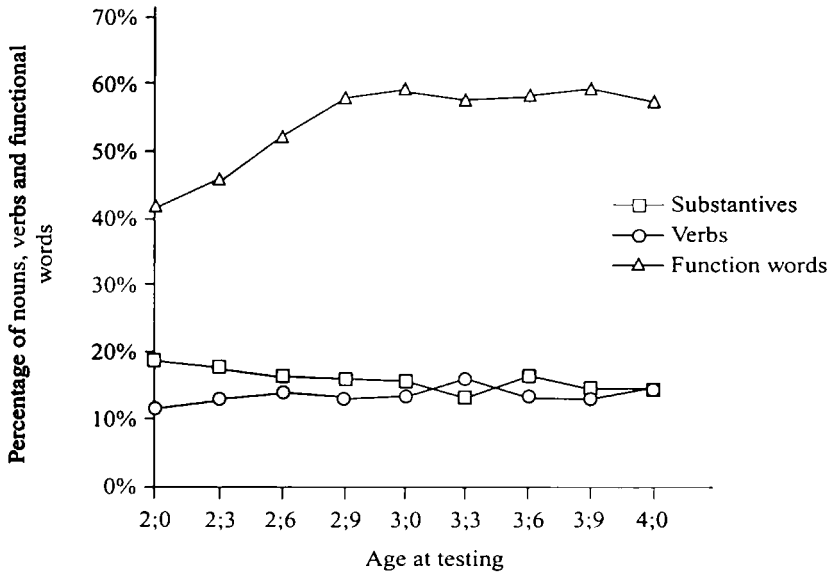
Fig. 2. **Proportion** of nouns, verbs and function words in **children's**
spontaneous production

categories are only just significant, but in this case the standard
deviation is already very small at the age of 2;0.

The evolution of local regularities is not only found in the global
variation between children. The variability between children also
diminishes for specific syntactic features such as the use of articles and
pronouns (see Fig. 1). This is particularly true for articles, where the
test for homogeneity of independent variances gives a result of
$F(26,27) = 38.08$, $p < 0.0001$. The result for pronouns is also highly
significant ($F(26,27) = 7.37$, $p < 0.0001$). The difference between child
and adult becomes very small by age 3;0, and declines still further
during the fourth year. At the age of four, there is no significant
difference between the use of articles in obligatory contexts by children
and adults: children produce an article with a noun 91% of the time,
adults 93% of the time ($t(36) = 1.76$, $p = 0.08$). The difference in the
use of a pronoun with a verb is greater, 82% for children and 87% for
adults, and is significant ($t(36) = 2.43$, $p = 0.02$).

Global regularities

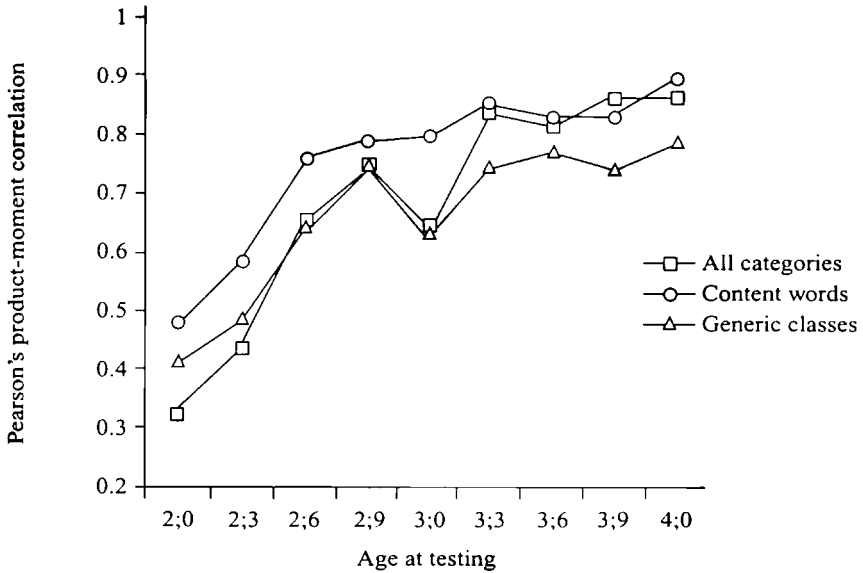The production of nouns, verbs and function words follows different

Fig. 3. The relationships between children's and adult's production of syntactic categories and content word classes

patterns, as seen in Fig. 2. The number of function words produced by children increases from 42% at age two to 57% at age four, and there is a significant difference between those two ages ($t(35) = 6.68$, $p < 0.0001$). The number of nouns decreases slightly but significantly, from 19% to 13% ($t(35) = 3.33$, $p = 0.002$), and the number of verbs remains the same, with no significant change ($t(35) = 1.35$, $p > 0.1$). The variability among children diminishes for nouns, verbs and function words. This is more evident for nouns, as shown with the test for homogeneity of independent variances ($F(26,27) = 9.50$, $p < 0.0001$). The same trend is seen for function words ($F(26,27) = 6.38$, $p < 0.0001$), and for verbs ($F(26,27) = 2.42$, $p = 0.03$).

The computation of the Pearson product-moment correlation between child and adult lexical category use – i.e., between one of the three first and the fourth columns of the Appendix – gives a global evaluation of the regularity of syntax use. Three correlation results are shown in Fig. 3: for all lexical categories (function words included), for all content word lexical categories (14 classes out of the full 25), and for seven generic classes of content words (nouns, verbs, adjectives, adverbs, communicators, interrogative-connector, object-focus). Whatever

the type of analysis performed, the global tendency is always the same: there is a steep increase in the correlation value between child and adult global production up to the age of 3;0, then development slows down during the fourth year. Over the same period, the standard deviation for the correlation value diminishes. At the age of 2;0, the standard deviation value for the correlation of all lexical categories is 0.18 for a mean of 0.32. At the age of 4;0, it is 0.09 for a mean of 0.86. The result for the test for homogeneity of independent variances is highly significant ($F(26,27) = 4.06, p = 0.0005$). The same type of result holds for content word lexical categories and generic classes of content words, suggesting a progression of local regularities, but with a much steeper upward curve. It should be noted that, although other global characteristics such as MLU and number of utterances produced increase with age, they show no reduction in inter-individual variability (see Table 1). On the contrary, the variability tends to increase with age.

DISCUSSION

The goal of this study was to show that different aspects of children's grammatical knowledge evolve at varying rates over the course of language acquisition. Another aim was to show that it is necessary to take into account quantitative data about children's production when one studies language acquisition.

The percentages obtained for syntactic category co-occurrences are higher than those obtained for words because of the part-of-speech generalization effect. For example, 'to think about something' could be generalized from 'to talk about something'. At the category level, both are correct if a single verb class is used to cover all types of verbs. Also, the adult corpus is not part of the input of the children studied as the two corpora were not recorded at the same time and with the same persons, so any lexical similarities only reflect the average lexicon produced by children.

Considering the number of French words on which this analysis was done, the results for tri-words cannot be simply random guesses and are even surprisingly high when compared with the bi-word results. If children merely produced adult words at random, there would be only one chance in 500 that they should produce a correct adult bi-word and one chance in 1,000,000 for a correct tri-word. The statistical difference between types and tokens is very low because this result corresponds to the mean of all children's percentages. Adults' collocations that are also found in child language are common to many of the children. Non-adult collocations are usually specific to individual

children, with the exception of utterances containing an exclamation (in French, *oh, ah*) which are very characteristic of and very common in children. When one considers each child individually, non-adult constructions are repeated as often as adult ones, so the difference between values for types and tokens is small.

Although the correct application of local morphosyntactic charac-teristics is confirmed, the degree of application evolves with age. This is not a new result, as morphosyntax gets progressively richer as children get older, but it is confirmed by our corpora. This can be seen in Fig. 1 where the evolution in the use of articles and pronouns in obligatory contexts follows a steep and regular curve towards the adult figures.

The findings of this study show that child language variation is more evident at a global level than at a local level. Clearly, there is a great increase in the correlations between syntactic category production by children and adults between the ages of 2;0 and 4;0. This is also true for the reduced set of content word categories. Conversely, as evidenced by the significant overlap between syntactic and lexical collocations produced by children and adults, adherence to morphosyntax appears early. Reliance on morphosyntax may start slowly, with very young children omitting many function words, but it is nonetheless applied correctly early on. Differences between children tend to disappear with age, and are rather small to begin with. Thus, the increase in children's consistency appears much steeper for the global features of language than for the local features.

Very often, variations among children are not clear-cut and are more visible at the earliest stages of language, in the second year (Bates & Marchman 1988, Bates, Marchman, Thal, Fenson, Dale, Reznick, Reilly & Hartung 1994, Bloom, Lightbown & Hood 1975, Lieven *et al.* 1992, Lieven, Pine & Baldwin 1997, Nelson 1973). Variability and regularity can be described and analysed in different ways. Local regularities show either that children closely follow the linguistic input they receive from adults and other children or, depending on the theory, that children have an instinctive mastery of morphosyntax. The fact that there is not much global regularity in content word categories before the age of four rather militates against the idea of an overall instinctive mastery of syntax and of generic classes such as verbs and nouns, as suggested by Pinker (1987). Pinker proposes that the link between the semantic and syntactic properties of language helps children in their early learning of grammar, i.e., the semantic bootstrap hypothesis. The understanding of semantic categories provides children with an understanding of the basic syntactic categories such as noun,

verb, adjective, preposition, sentence, etc. (Pinker 1987). It is not the semantic bootstrap principle which is criticized here, but the grammar to which it is hypothesized to be connected or the developmental path this connection follows. According to strong hypotheses about the status of noun and verb, global regularities should appear sooner than they do, at least by age three, if not before. In classic linguistic theories such as those of Pinker (1984, 1987) and Radford (1990), the local morphosyntactic properties of a language are usually considered secondary to fundamental lexico-semantic categories such as noun and verb. Thus, in these theories, if there is no mastery of innate lexico-semantic categories at age two, it follows that there is no mastery of innate local morphosyntactic properties. Based on the results in the present study, however, it seems that the opposite holds true: local regularities show that when children pick up chunks of input, they do not stop at adults' word boundaries but reproduce the morphosyntactic properties of the language they are learning – even though they do not reproduce all the morphosyntax. In French, for example, determiners and pronouns are produced very early, as these categories are always in close proximity to content words.

The local mechanism of morphosyntax cannot provide regularities at a sentence or text level. If there is no global structure in the utterances produced – and thus a highly variable proportion of nouns, verbs, adjectives, etc. – global regularities cannot possibly exist, even though the child may have a very adequate mastery of morphosyntax. Knowledge of morphosyntax may explain why the evolution for all word classes follows the curve for generic classes, as shown in Fig. 3. It is around generic classes that morphosyntax is built. Regularity in generic classes is necessary for global regularity in all classes.

For two-year-olds, global irregularities may simply express the randomness of children's input and their incomplete cognitive mastery. It cannot be a mere reflection of the randomness of adult syntax, because adult output is certainly syntactically uniform. But it may reflect the randomness of the child-adult interaction as well as the child's understanding of the world. If a child was reproducing only phonetic patterns, his or her production would contain as many function words as content words, especially in languages such as French where the phonetics of function words is not much different from that of content words. Ruling out phonetics and syntax, only cognitive mechanisms can explain the high production of nouns and of content words. However, cognition and syntax are not necessarily always fully correlated, as demonstrated by Maratsos (1982).

The mastery of the local morphosyntactic properties of a language,

which occurs before or in parallel to the mastery of more syntactic-cognitive categories, should participate actively in determining the properties of the grammar being learned. At the age of three, a child's understanding of cognitive categories is still partial, explaining the continuing improvement in global linguistic mechanisms between the ages of three and four. It still remains to be shown how local morphosyntactic mechanisms suffice to give a 'correct' appearance to children's short utterances, and how adults' automatic linguistic mechanisms are rooted in the coupling of morphosyntactic rules and the cognitive and semantic knowledge of the world.

A second contribution of this article was to show that quantitative analysis of language transcripts can offer new insights into language development. For example, Radford (1990) emphasizes that his study was conducted using naturalistic samples and a large database of nearly 100,000 utterances. Using naturalistic data and a large corpus of examples is certainly a very good practice, but Radford's samples were without reference to frequency, repetitiveness, and other measures that help to assess the import of the structures described. When using knowledge of children's production in building a theory, one must keep in mind how often and in what circumstances the production occurs, be it normal or erroneous. The dynamics of the production figures is also very important. On the other hand, it is not good to use only statistical data. There are many reasons why looking at the corpus and finding concrete examples is necessary and why large-scale statistics cannot provide all the necessary information about language development. For example, children's errors need to be examined closely, especially when they are infrequent. Thus, an amalgamation of the classical approach to child language and the approach used here is probably a requisite to further progress in our understanding of how children learn and acquire language.

REFERENCES

Bates, E. & Marchman, V. A. (1988). What is and is not universal in language acquisition. In F. Plum (ed.), *Language, Communication, and the Brain* (New York: Raven Press.).

Bates, E., Marchman, V. A., Thal, D., Fenson, L., Dale, P. S., Reznick, J. S., Reilly, J. S. & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, 21, 85–123.

Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars* (Cambridge, MA: MIT Press).

Bloom, L., Lightbown, P. M. & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, 402, serial no. 160.

Bowerman, M. F. (1973). *Early Syntactic Development: A Crosslinguistic Study with Special Reference to Finnish* (London: Cambridge University Press).

Brown, R. W. (1973). *A First Language: The Early Stages* (Cambridge, MA: Harvard University Press).

Caselli, C., Casadio, P. & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*, 26, 69–111.

Charniak, E. (1993). *Statistical Language Learning* (Cambridge, MA: MIT Press).

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

Kemper, S., Rice, K. & Chen, Y. J. (1995). Complexity metrics and growth curves for measuring grammatical development from five to ten. *First Language*, 15, 151–66.

Le Normand, M. T. (1986). A developmental exploration of language used to accompany symbolic play in young, normal children (2–4 years old). *Child, Care, Health and Development*, 12, 121–34.

Lieven, E. V., Pine, J. M. & Baldwin, W. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187–219.

Lieven, E. V. M., Pine, J. M. & Barnes, H. D. (1992). Individual differences in early vocabulary development: redefining the referential-expressive distinction. *Journal of Child Language*, 19, 287–310.

Lowe, M. & Costello, A. J. (1976). *Manual for Symbolic Play Test, Experimental Edition* (Windsor: NFER-Nelson Publishing Company).

MacWhinney, B. (1978). The acquisition of morphology. *Monographs of the Society for Research in Child Development*, 43 (1–2), serial no. 174.

—— (1995). *The CHILDES Project: Tools for Analyzing Talk* (Hillsdale, NJ: Lawrence Erlbaum).

MacWhinney, B. & Snow, C. E. (1985). The child language data exchange system. *Journal of Child Language*, 12, 271–96.

Maratsos, M. P. (1982). The child's construction of grammatical categories. In E. Wanner & L. R. Gleitman (eds), *Language Acquisition – the State of the Art* (New York: Cambridge University Press).

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

Parisse, C. & Le Normand, M. T. (1998). Traitement automatique de la morphosyntaxe chez le petit enfant. *Glossa*, 61, 22–9.

—— (2000). How children build their morphosyntax: the case of French. *Journal of Child Language*, 27, 267–92.

Pinker, S. (1984). *Language Learnability and Language Development* (Cambridge, MA: Harvard University Press).

—— (1987). The bootstrapping problems in language acquisition. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition* (New York: Springer-Verlag.).

Radford, A. (1990). *Syntactic Theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English* (Oxford: Basil Blackwell).

Robinson, B. F. & Mervis, C. B. (1998). Disentangling early language development: modeling lexical and grammatical acquisition using and extension of case-study methodology. *Developmental Psychology*, 34, 363–75.

Rondal, J. A. B. J. F. & Pérée, F. (1985). Analyse du langage et des interactions verbales adulte-enfant. *Bulletin d'Audiophonologie*, 5, 507–36.

Suppes, P., Léveillé, M. & Smith, R. (1974). *Developmental Models of Child's French Syntax* (*Technical Report # 243*, Stanford University).

Suppes, P., Smith, R. & Léveillé, M. (1972). *The French Syntax and Semantics of Philippe, Part 1: Noun Phrases* (*Technical report # 195*, Stanford University).

Wells, G. (1985). *Language Development in the Preschool Years* (New York: Cambridge University Press).

APPENDIX

List of 25 morphosyntactic categories produced by children aged 2;0, 3;0 and 4;0, and by adults

Class tag	Percentage of use				Description of the morphosyntactic class
	at 2;0	at 3;0	at 4;0	Adults	
adj	3	3	3	3	Adjective
adv	4	4	6	6	Adverb
adv:neg	3	4	3	4	Adverb of negation
adv:place	7	3	2	1	Adverb of place
adv:voici	3	1	0	0	Locution *voici*, *voilà*
co	8	1	2	2	Communicator
co:act	7	1	1	1	Exclamation
conj	1	2	4	4	Conjunction
det	5	10	9	7	Article
det:gen	0	1	2	2	Generalized article
n	18	14	13	11	Noun
n:prop	4	1	1	2	Last name, proper noun
num	0	1	1	0	Number
prep	0	4	5	5	Preposition
prep:art	1	1	1	2	Preposition article
pro	8	17	17	18	Pronoun
pro:dem	3	3	2	2	Demonstrative pronoun
pro:rel	2	2	3	6	Relative or interrogative pronoun
pro:y	1	2	2	1	Pronouns *y*, *en*
v	3	5	5	7	Verb
v:aux|avoir	2	3	3	3	Verb *avoir* (to have)
v:aux|être	7	5	4	4	Verb *être* (to be)
v:inf	3	5	4	3	Infinitive
v:mdl	1	4	4	3	Modal
v:pp	5	2	3	3	Past participle
v:prog	0	0	0	0	Present progressive