A comparison of the transition from first words to grammar in English and Italian*

CRISTINA CASELLI

Institute of Psychology, National Council of Research, Rome

PAOLA CASADIO

Santa Lucia Neurological Clinic and Research Center, Rome

ELIZABETH BATES

University of California, San Diego

ABSTRACT

Cross-linguistic similarities and differences in early lexical and grammatical development are reported for 1001 English-speaking children and 386 Italian-speaking children between 1;6 and 2;6. Parents completed the English or Italian versions of the MacArthur Communicative Development Inventory: Words and Sentences, a parent report instrument that provides information about vocabulary size, vocabulary composition and grammatical complexity across this age range. The onset and subsequent growth of nouns, predicates, function words and social terms proved to be quite similar in both languages. No support was found for the prediction that verbs would emerge earlier in Italian, although Italians did produce a higher proportion of social terms, and there were small but intriguing differences in the shape of the growth curve for grammatical function words. A strikingly similar nonlinear relationship between grammatical complexity and vocabulary size was observed in both languages, and examination of the order in which function words are acquired also yielded more similarities than differences. However, a comparison of the longest sentences reported for a subset of children demonstrates large cross-linguistic differences in the

amount of morphology that has been acquired in children matched for vocabulary size. Discussion revolves around the interplay between language-specific variations in the input to young children, and universal cognitive and social constraints on language development.

INTRODUCTION

A primary goal of developmental psycholinguistics is to uncover the universal mechanisms that govern language development. Cross-linguistic studies have played a major role in that effort. By moving outside the boundaries of a single language (e.g. English), we are able to disentangle the confound between universal mechanisms and language-specific content, while exploring the wide range of variations that can be observed in the language-learning process. In the last 25 years, there has been a marked increase in cross-linguistic research on language development, at every level from speech perception in infancy (Kuhl, 1991; Goodman & Nusbaum, 1994; Werker, 1994), through the acquisition of grammar (Slobin, 1985–97; see also MacWhinney & Bates, 1989), to the finer details of narrative discourse from preschool through the elementary school years (Berman & Slobin, 1994). At all these levels, the evidence points to cross-linguistic variations in content (phonetic, lexical and/or grammatical), complemented by evidence for constraints on learning, perception and production that operate in every language.

In this paper, we will investigate this interplay in a brief but very important period of development from 1;6 to 2;6, when children make the passage from first words to grammar. We will compare this transition in English and Italian, languages that provide an interesting test of two related issues: (1) universal vs. language-specific patterns in the composition of vocabulary, with special emphasis on the onset and growth of verbs and function words as a function of total vocabulary size, (2) the relationship between vocabulary size and grammatical complexity, including qualitative evidence about the order in which specific function words emerge and the amount of grammatical morphology that we see in each language when children are matched for vocabulary size. As we shall see, these two languages provide strong evidence for universal constraints on developmental changes in the composition of vocabulary and on the overall relationship between grammar and the lexicon, although subtle cross-language variations in lexical and grammatical content are observed.

Before presenting results of a study focused on those issues, a brief review of the current literature on both points is in order.

Cross-linguistic similarities and differences in lexical development

Language-specific variations in lexical content are inevitable in early development, reflecting the arbitrary relationship between sound and meaning

across languages (e.g. 'dog' in English, 'cane' in Italian), together with differences in the statistical distribution of comparable word types as a function of cultural factors (e.g. cross-linguistic differences in the relative frequency of words like 'spaghetti'). Nevertheless, some possible universal stages in the composition of early vocabulary have been proposed, hypothesized to reflect universal cognitive and social constraints that override language-specific variations in content. Perhaps the best-known proposal of this kind comes from Gentner (1982), who has argued that verbs must develop later than nouns in all human languages (more on this below). Based on our own findings for English (Bates, Bretherton & Snyder, 1988; Bates, Marchman, Thal, Fenson, Dale, Reznick, Reilly & Hartung, 1994) and Italian (Bates, Benigni, Bretherton, Camaioni & Volterra, 1979; Caselli, Bates, Casadio, Fenson, Fenson, Sanderl & Weir, 1995), we have expanded Gentner's noun–verb proposal into a four-stage model of lexical development that includes hypotheses about lexical content before and after the noun–verb transition, as follows:

Routines and word games. In the very first phase of lexical development, when expressive vocabularies range from 0 to 10 words, children tend to produce words that are difficult if not impossible to classify in adult part-of-speech categories, including sound effects for animals and vehicles, social routines like 'bye', 'hi' and 'uh-oh', and names for favourite people. These verbal routines are best viewed as SPEECH ACTS or PERFORMATIVES, vocal conventions that children use in familiar and well-structured situations to achieve some social function (see also Nelson & Luciarello, 1985; Dromi, 1987; Lieven & Pine, 1990; Ninio, 1993). In fact, categories like 'noun' and 'verb' may not be operating at all in this early phase of development (Tomasello, 1992).

Reference. When expressive vocabulary grows to between 50 and 200 words, the overwhelming majority of words are nominals (broadly defined). Even when we restrict the definition of nominals to common nouns (i.e. names for classes of concrete objects), it is difficult to escape the conclusion that nouns predominate and grow sharply, in absolute numbers and as a proportion of all word types. Although there are individual differences along this dimension (Nelson, 1973), and some children still engage in a large number of routines and word games that defy classification in adult terms, for most children this period of development revolves primarily around words that establish reference.

Predication. Verbs and adjectives are very rare in the first two periods of lexical development, comprising between 0 and 5 % of all words for most English-speaking children. These categories undergo a notable increase after the first 100 words, in absolute numbers and as a proportion of all word types. It has been argued that this change in vocabulary composition reflects the emergence of predication, i.e. the ability to encode relational meanings.

71

Indeed, it is probably no coincidence that word combinations do not appear before the 50-word point (Nelson, 1973), and are not produced consistently until children achieve vocabularies between 100 and 200 words (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994).

Grammar. Grammatical function words are also extremely rare in the first stages of lexical development. Bates *et al*. (1994) report that these terms constitute less than 5 % of all words in the first and second year of life, and do not display proportional growth until children achieve a total expressive vocabulary between 300 and 500 words. The occurrence of function words prior to the 400-word point appears to be uncorrelated or negatively correlated with measures of grammatical development after that point, suggesting that the first function words are learned as memorized routines that may bear little relationship to the emergence of productive grammar. By contrast, the proportional growth of function words after the 400-word point coincides and correlates with various indices of grammatical productivity, including mean length of utterance in morphemes and alternative measures of inflectional productivity.

These changes in the composition of vocabulary are hypothesized to reflect universal developments in the logical and conceptual substrates of meaning (O'Grady, 1987). A compelling rationale for at least one part of this story appears in an influential paper by Gentner (1982), who argued that concrete nouns MUST precede verbs in early language development because nouns are easy to grasp (based on concrete objects that 'hold still' long enough to support word learning), while verbs reflect relational meanings that are harder to perceive and (more important) defined by a network of meanings that are subject to language-specific and situation-specific variations (e.g. the difference between 'give' and 'put', which depends on the animate or inanimate nature of the dative object). However, this proposed universal has been challenged recently in cross-linguistic studies of Korea (Gopnik & Choi, 1990, 1995; Choi & Gopnik, 1995) and Chinese (Cheng, 1994; Tardif, 1996; Tardif, Shatz & Naigles, 1996), with implications for the whole chain of events leading up to grammar that we have just described. In particular, it has been proposed that children who are learning Korean or Chinese do not display the same early bias toward nouns that has been observed in children learning English (e.g. Gopnik & Choi, 1995; Tardif, 1996). These investigators argue that the proposed universal transition from nouns to verbs may be an epiphenomenon of the fact that most studies of early lexical development have been based on English. They also note that most studies reporting an early noun advantage have been based on parental report and/or on data collection in cultural contexts that emphasize object naming. By contrast, data for children learning Korean or Chinese suggest that verbs may be far more common in these languages, even within the very first stages of lexical learning. Indeed, Tardif claims that verbs may actually predominate

72

statistically over nouns in many Chinese children, a direct challenge to the hypothesized universal-stage model described above.

To explain their findings for Korean, Gopnik and Choi point out a number of differences between this language and the others studied to date which could mitigate against a noun advantage. For example, Korean is an SOV (subject–object–verb) language, which means that verbs are usually in a salient position. Korean also makes extensive use of both subject and object omission, which means that verbs are often the only content word in sentences spoken to young children. In addition to these differences in linguistic structure, the authors point out some potentially relevant cultural differences between English-speaking and Korean-speaking mothers, re-volving around the relative emphasis on object play (with object naming) and other forms of social interaction. Arguments in the same spirit are offered by Tardif for Chinese. Unlike Korean and like English, Chinese is pre-dominantly an SVO language. However, Chinese permits far more word order variation than English, and (like Korea) Chinese also permits extensive omission of both the subject and the object. In addition, Chinese has no grammatical inflections of any kind that might be used to distinguish between nouns and verbs, a fact which may make it easier for children to cross the boundary between adult form classes at an early age.

These challenges to the proposed universal sequence of lexical de-velopment have been controversial, but they are also plausible and important, warranting more extensive investigation. In an earlier cross-linguistic study of lexical development between 0;8 and 1;4 (Caselli, Bates, Casadio, Fenson, Fenson, Sanderl & Weir, 1995), we suggested that Italian might provide a good test case for the early onset of verbs. For example, although Italian is an SVO language, it permits extensive word order variation. It is also a prodrop language in which subjects are omitted around 70% of the time in informal conversation (Bates, 1976). Because of these characteristics, verbs are often located in sentence-initial or sentence-final positions that are easy for children to perceive, a situation analogous to the one described for Korean and Chinese. Unlike these two languages, however, Italian has an extremely rich system of verb morphology, and verb agreement plays a crucial role in conveying basic sentence relations. Indeed, current evidence suggests that Italian children are sensitive to verb agreement from a very early age, in both comprehension and production (Pizzuto & Caselli, 1992, 1993; Devescovi, D'Amico, Smith, Mimica & Bates, in press). Taken together, these features of Italian would appear to provide a solid basis for cross-language variation in verb onset and verb growth.

Like the study that we will present below, the Caselli *et al*. study was based on a parental-report instrument called the MacArthur Communicative Development Inventory, or CDI. The CDI contains two separate scales: the Words and Gestures Scale, designed to measure word comprehension, word

production and gesture between 0;8 and 1;4, and the Words and Phrases Scale, used to assess word production and various aspects of grammar between 1;4 and 2;6. Caselli *et al.* used the Words and Gestures inventory to assess lexical development in 195 Italian infants between 0;8 and 1;4, compared with the 659 American infants in the same age range from the MacArthur CDI norming study (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick & Reilly, 1993; Fenson *et al.*, 1994). Findings were disappointing for the relativist perspective outlined by Gopnik, Choi and Tardif, providing evidence in favour of the universalist account proposed by Gentner and colleagues (1982, 1997; Gentner & Boroditsky, in press). In Italian as well as English, nouns overwhelmed verbs in onset, number, and rate of growth throughout this age range, in absolute terms and as a proportion of total vocabulary.

One might argue that comparisons based on this parental report form provide an unfair test of cross-linguistic differences. In particular, because the CDI was originally developed for English, adaptations of that instrument to other languages may be strongly biased toward English. There are two reasons why we believe that this criticism does not apply here. First, all non-English adaptations of the CDI are true adaptations, not translations of the English scales. Items in the vocabulary checklist (and the grammatical subscales – see below) are drawn from the existing literature on early development in that language, selected to reflect the lexical and grammatical forms that are known to appear in the relevant windows of development (Bottari, Cipriani, Pfanner & Chilosi, 1993; Cipriani, Chilosi, Bottari & Pfanner, 1993; see also Jackson-Maldonado, Thal, Marchman, Bates & Gutierrez-Clellen, 1993). They are also pre-tested with parents who are native speakers of the language, seeking their advice about items that should be added or dropped. Secondly, there is a unique historical relationship between the English and Italian versions of the CDI. Starting with a joint study in the 1970s of English and Italian infants between 0;9 and 1;1 (Bates *et al.*, 1979), the respective English and Italian instruments have been developed in parallel, in a series of joint and independent projects across a 20-year period. Hence the English version did not precede the Italian one in any relevant sense. To be sure, the current scales are designed to permit cross-linguistic comparisons, holding the age appropriateness and number of items within each subscale constant (see Method, below). However, the content of both the vocabulary and the grammar scales is determined by the structure and the relevant developmental facts of each language.

Caselli *et al.* discuss other methodological differences between their findings for Italian and contrasting findings for Chinese and Korean that could explain the apparent contradictions in our results. Because these points are relevant to the cross-linguistic test with older children that we will present below, they are summarized briefly here.

The most obvious contrast lies in the fact that these studies are looking at different languages. Perhaps, one might argue, our assumptions that verbs ought to be more salient in Italian is simply wrong. While we cannot rule this out, we are perplexed, because most of the relevant conditions listed by Gopnik, Choi and Tardif also hold for Italian, e.g. differences in salience due to verb position and subject omission. We are thus tempted to conclude that the most important methodological difference between our studies and those that have been reported for Chinese and Korean are due to the methods used to assess language production. As reviewed in some detail by Au, Dapretto & Song (1994), Tardif (1996), Gentner (1997), Gentner & Boroditsky (in press), those studies that have been successful in uncovering a verb advantage have relied entirely on free speech samples and/or on short parental checklists that were specifically designed to pick up information about non-nominal expressions. When Pae (1993) and Au *et al*. (1994) conducted independent studies of lexical development in Korean using parent report checklists similar in length and representativeness to the ones that we have used here, they obtained results for Korean that were very close to our findings for Italian (e.g. nouns outnumber verbs at every point). Further-more, Au *et al*. also conducted an experimental study in which American and Korean children were presented with novel nouns and verbs; in both samples, children found it much easier to learn new object names, compared with names for a novel action. Similar studies have not been conducted for Chinese, so it remains possible that future work will uncover a verb advantage in this language using parent report and/or novel word learning. If this turns out to be the case, we may have to reconsider the structural differences that separate Chinese from the other languages studied to date.

Nevertheless, we would still need to account for the fact that free speech and parent report yield very different results within at least one language, Korean. Putting together the studies by Gopnik and Choi on the one hand, and Pae and Au *et al*. on the other, we suggest that there are strong universal constraints on the development of *lexical knowledge*, but these universals coexist with language-specific profiles of *lexical use*. The most relevant comparison in this regard comes from Tardif *et al*. (1996), who compared free speech for small samples of English, Chinese and Italian children in the age range of 1;8 to 1;10. In that study, English-learning children produced (as always) far more nouns than verbs. Italian children also produced more nouns than verbs, although they showed a slight verb advantage compared to the American sample. By contrast, the Chinese children almost always produced more verbs than nouns. It thus seems plausible to conclude that there are cross-linguistic differences in the forms that children like to use in free speech settings during the early stages of language development.

Perhaps this should not surprise us, in view of differences between knowledge and use that have been reported in previous studies of English.

75

For example, Bates *et al.* (1988) compared the percentage of verbs over all word types in parent report vs. free speech, and found remarkably low correlations between these two measures. Furthermore, the free speech verb proportion scores correlated with a very different set of variables than the parent report verb proportion scores, leading these authors to conclude that these two measures are tapping into very different abilities. To account for these differences between methodologies, Bates *et al.* propose that free speech tells about WHAT THE CHILD LIKES TO DO, while parent report tells us more about WHAT THE CHILD KNOWS. Based on these and other findings, we have proposed a related hypothesis for cross-linguistic research: CROSS-LINGUISTIC DIFFERENCES IN LEXICAL AND GRAMMATICAL DEVELOPMENT WILL BE GREATER IN STUDIES USING METHODS THAT ARE SENSITIVE TO THE STATISTICAL PROBABILITIES AND CONTEXTUAL PREFERENCES THAT CHARACTERIZE EVERYDAY LANGUAGE USE.

In addition to these differences in methods of data collection, the above studies also vary in the age range under test. The results for Chinese and Korean that we have described so far are based primarily on studies of children who are older than 1;9; by contrast, Caselli *et al.* looked at vocabulary composition and growth from 0;8 to 1;4, in the very earliest phase of lexical development. It is possible that the relevant cross-language differences in vocabulary composition do not appear until some point after 1;9, when children start to understand and acquire the structural contrasts that we outlined above (e.g. subject omission in Italian, obligatory subjects in English; word order variation in Italian, preservation of SVO in English). Hence, in the present study, we will compare developmental changes in the composition of vocabulary in English and Italian children between 1:6 and 2;6.

Cross-linguistic similarities and differences in the relation between grammar and the lexicon

A second goal of the present study is to provide a cross-linguistic test of the powerful relationship between vocabulary size and grammatical development that has been reported for children acquiring English (Bates *et al.*, 1994; Fenson *et al.*, 1994; Marchman & Bates, 1994; Bates & Goodman, 1997). This includes the proportional increase in number of function words described above for children with vocabularies over 400 words, but it also includes a tight nonlinear correlation between vocabulary size and sentence complexity. Bates & Goodman argue that the strong interdependence of grammar and the lexicon during this period of development provides evidence in favour of lexicalist theories in which the development of vocabulary and grammar are based on common mechanisms, and against theories in which grammar is an autonomous module that is structurally and developmentally separate from the lexicon. Furthermore, Bates and Good-

76

man propose that this nonlinear relationship between vocabulary size and grammar may be a universal property of language development – a strong hypothesis that merits a rigorous cross-linguistic test. In the present study, we will compare the relationship between lexical and grammatical development in a large sample of English and Italian children, during the period from 1;6 to 2;6, when most normal children complete the transition from single-word utterances to productive control over grammar. Analyses will focus on two aspects of grammar: overall grammatical complexity, and the emergence of specific function words. We will also provide a brief qualitative look at the relationship between vocabulary size and inflectional morphology in a handful of cases from each language matched for vocabulary size, to illustrate the fact that the content of grammar is quite different in these languages, despite the similarities in growth and sequencing that we see for comparable structures.

It is important to place this approach to early grammar into its historical perspective. Thanks in large measure to pioneering research by Dan Slobin and his international network of collaborators (Slobin (ed.), 1985–97), there is now an ample cross-linguistic literature on early grammatical development. This comparative approach has also been extended fruitfully to the development of narrative discourse from preschool through elementary school years, with an emphasis on cross-language variations in the relationship between discourse functions and the specific lexical and grammatical devices used to realize those functions (Karmiloff-Smith, 1979; Bamberg, 1988; Berman & Slobin, 1994). In all these studies, the evidence for cross-language variation has been striking. For example, we now know that the onset and growth of inflectional morphology can vary markedly from one language to another, starting as early as the one-word stage in some richly inflected languages. There are also dramatic cross-language contrasts in the word orders that predominate in first word combinations, and in the degree to which word order regularities are observed at all (Bates, 1976; Braine, 1976). The appearance of complex syntactic structures is conditioned by cross-language variations in the input. For example, passives appear as early as 2;0 in Sesotho, a language in which passives are very frequent (Demuth, 1989). Similarly, relative clauses are far more common in the speech of Italian than they are in English children at 3;0, reflecting the differential frequencies of the relative clause in English and Italian adults (Bates & Devescovi, 1989). Finally, studies of narrative discourse by Berman, Slobin and colleagues provide evidence for cross-linguistic contrasts in those aspects of the situation that children choose to encode, and in the lexical and grammatical devices that they select to convey the same set of discourse functions (e.g. tense, aspect, foregrounding, backgrounding). Despite all this evidence for cross-language variation, children in every language community commit the same kinds of errors (e.g. overgeneralizations), and they show preferences that can

77

be characterized in universal terms (e.g. 'Pay attention to the ends of words', 'Avoid discontinuous elements' – Slobin, 1985). These trends suggest that grammatical development is subject to universal constraints, superimposed on input-driven variation.

In contrast with this rich comparative literature on syntax and morphology, relatively little is known about early developments at the interface between grammar and vocabulary, including differences in the acquisition of grammatical function words. There are, of course, some interesting and instructive exceptions. For example, Slobin's early cross-linguistic projects placed great emphasis on universal and language-specific patterns in the acquisition of locatives (including locative markers on nouns, and locative prepositions, e.g. Johnston, 1985). As we might expect, these comparisons provided evidence for universal cognitive constraints (e.g. locatives that express complex spatial relations like 'in front of' and 'behind' emerge later than locatives that encode simpler relations like 'in' and 'on'). At the same time, it is clear from this early work (and from more recent studies of locatives in English and Korean – Choi & Bowerman, 1991; McDonough, Choi, Bowerman & Mandler, in press) that the child's linguistic input can have an influence in the order in which locative prepositions are acquired, and in the specific way that space is carved up for linguistic expression. Similar cross-linguistic findings have been reported for linguistic terms that mark the relations between space and time (Weist, Lyytinen, Wysocka & Atanassova, 1997) and the distinction between count and mass nouns (Gathercole & Min, 1997; Imai & Gentner, 1997).

In the present study, we have a unique opportunity to compare aspects of early grammar in unusually large samples of English and Italian-speaking children, controlling for age and for changes in vocabulary size. Of course there are clear limitations on our ability to investigate the details of grammar using a parent report technique. Such procedures will never replace traditional free-speech and/or experimental measures in advancing our knowledge of grammatical development. Nevertheless, we can learn something about gross changes in structural complexity, and about the relative onset and growth of different function words, providing some working hypotheses for more focused observational and experimental studies.

Finally, before proceeding, we should clarify that the present study compares results for two large data bases that have been described separately in other published works. The English data come from a large cross-sectional sample collected by Fenson et al. (1993, 1994), described in the first published norms for the MacArthur Communicative Development Inventories. The Italian data come from another large norming sample collected and described by Caselli & Casadio (1995), in published norms for the Italian version of the MacArthur. For the English sample, Bates et al. (1994) have already described developmental changes in the composition of

vocabulary; comparable analyses for the Italian 1;6–2;6 sample have not been published before. Bates & Goodman (1997) have described the relation between grammatical complexity and vocabulary size in the English sample, and in a separate longitudinal sample of English-speaking children. Comparable analyses of the Italian sample have not been published before, although Bates & Goodman do make brief reference to our (then) unpublished findings for Italian. Finally, as noted above, Caselli *et al.* (1995) brought the two norming samples together for the first time, in analyses restricted to the period between 0;8 and 1;4. The present study is a sequel to Caselli *et al.*, presenting (also for the first time) a systematic cross-linguistic comparison of growth in vocabulary and grammar, changes in vocabulary composition, and relations between grammar and vocabulary in English and Italian children between 1;6 and 2;6.

METHOD

Subjects

The English-speaking sample for the present study comprises 1001 children between 1;6 and 2;6, from the norming sample described by Fenson *et al.* (1993), Fenson *et al.* (1994) and Caselli *et al.* (1995). Data were collected at three sites (San Diego, Seattle and New Haven), and procedures for contacting families differed slightly at those sites (see Fenson *et al.*, 1993, for details). An effort was made to obtain as representative a sample as possible of the ethnic, educational and social class characteristics that characterize these three cities, although the final sample was (as is so often the case in developmental research) skewed in the direction of educated, middle-class families. Parents who completed the MacArthur CDI were also asked to complete a 'basic information form', which supplied data about their child's health history and their own education, occupation and other pertinent information. Children were excluded from the study on the basis of several criteria derived from the basic information form, including history of mental retardation, prematurity (six or more weeks premature), extended surgical procedures or any other serious medical complications. The sample was also limited to families for whom English is the primary language used in the home.

   The Italian sample studied here comes from the norming study for the Italian CDI described by Caselli *et al.*; further details about the recruitment procedure and the demographic characteristics of the sample are available in Caselli & Casadio (1995). The sample includes 386 children between 1;6 and 2;6. Families from several different Italian cities participated in the study, although the majority come from the northern and central regions of the country. Although an effort was made to obtain a sample that is representative

of the Italian population as a whole, the education and social class characteristics of the final sample are also skewed toward middle-class families. Parents in the Italian study filled out a basic information form similar to the one adopted by Fenson *et al.*, and children were excluded for the same biomedical risk factors described above. All children came from families in which Italian is the primary language spoken in the home.

Table 1 summarizes the distribution of these samples by language, age and gender. Non-parametric statistics conducted within each language indicated that age and gender were unconfounded, and there were also no statistically reliable confounds in the gender distribution for English vs. Italian. Table 1

TABLE 1. *Breakdown of the two samples by age and gender*

| | English | | Italian | |
Age	Male	Female	Male	Female
1;6	46	34	8	7
1;7	35	37	14	11
1;8	48	45	27	25
1;9	40	31	15	17
1;10	32	40	21	20
1;11	41	41	13	19
2;0	48	59	21	20
2;1	31	45	16	11
2;2	39	39	8	10
2;3	42	38	13	9
2;4	30	31	17	7
2;5	33	28	15	10
2;6	32	36	15	17

shows that the smaller Italian sample is less evenly distributed across age levels than its larger English counterpart, and a likelihood ratio comparing the distribution over language and age does reach significance ($p < 0.02$). However, the two languages do not differ in the relative representation of younger and older children. In both groups, there are slightly fewer children represented in the later age groups; within each group, the line of best fit for the relationship between age and sample size has a slope between -0.10 and -0.11, indicating that this trend is in the same direction for both languages even though the variance is greater in Italian.

Materials

Data for the English sample are based on the CDI: Words and Sentences (Fenson *et al.*, 1993, 1994), designed for use with children in the age range from 1;4 to 2;6. This scale includes a 680-word vocabulary production checklist, organized into 22 semantic categories. The checklist is based on a

number of antecedent questionnaires, developed over a 20-year period. The earliest versions of the checklist were based on surveys of the literature (diary studies as well as naturalistic observations); words were added or dropped in successive studies in response to parental feedback. Vocabulary estimates from the CDI correlate highly with number of word types and tokens in studies of the same children, and the scales also have high internal consistency and test-retest reliability (for details, see Fenson *et al.*, 1993, 1994).

Data for the Italian sample are based on the Words and Sentences Scale for the Italian adaptation of the MacArthur CDI (Caselli & Casadio, 1995). For the Italian version of this Scale, norms are available between 1;6 and 2;6. The Italian word production checklist contains 670 items, organized into 23 semantic categories. The English and Italian checklists have the same number of open-class (content word) categories, and the same number of items within each of these categories, even though the specific items differ between languages in culturally and linguistically appropriate ways (e.g. different kinds of foods, clothing, etc.). In both languages, the open-class categories are 'sound effects' (12 items), 'animals' (43 items), 'toys' (14 items), 'food and drink' (68 items), 'clothing' (28 items), 'body parts' (27 items), 'small household objects' (50 items), 'furniture and rooms' (33 items), 'places to go' (22 items), 'names for people' (29 items), 'games and routines' (25 items), verbs (called 'action words', 103 items), adjectives and qualities (called 'descriptive words', 63 items). The closed-class or grammatical function word categories are similar in many respects in English and Italian, although they necessarily differ in both size and content, reflecting differences in the structure of these two languages and in their respective literatures on early language acquisition. Both languages have items under 'time words' and 'places to go' that were excluded from both the open- and closed-class analyses because so many of the items are ambiguous with regard to the open/closed distinction. The remaining items all counted as closed-class words in the analyses presented below (a total of 92 items for Italian and 102 for English), including 'pronouns' (25 items in English, 23 in Italian), 'question words' (seven each in English and Italian), 'prepositions' (26 in English, 17 in Italian), 'articles and quantifiers' (17 in English, 21 in Italian), auxiliary and modal verbs (called 'helping verbs' – 21 in English, 15 in Italian), and 'connecting words' (six each in English and Italian). In addition, the Italian checklist has a separate category called 'adverbials', with three additional closed-class words. For the analyses presented below, these three items ('ecco', 'qua/qui' and 'la/li') are grouped with the preposition and locative class. A complete list of all function word items, in each language, is provided later (see Results).

The English and Italian versions of the CDI contain several different subscales designed to measure aspects of grammatical development. However, as Fenson *et al.* (1994) have shown for English and Caselli & Casadio

have shown for Italian, these scales are highly intercorrelated. For our purposes here, we will therefore concentrate primarily on the grammatical complexity checklist for each language. Each grammar checklist comprises 37 sentence pairs, each reflecting a single morphological or syntactic contrast that is known to emerge across the period from 1;6 to 2;6, including developments in bound morphology (e.g. 'Daddy car' vs. 'Daddy's car'), presence or absence of obligatory function words ('I like read stories' vs. 'I like to read stories'; 'No wash dolly' vs. 'Don't wash dolly'), and variations in syntactic complexity through addition of non-obligatory elements ('Baby crying' vs. 'Baby crying cuz she's sad'; 'Want cookies' vs. 'Want cookies and milk'). Parents were asked to indicate (even if their child had not said this particular sentence) which sentence in each pair 'sounds more like the way that your child is talking right now.' Within each pair, the second alternative always represents a 'higher' (more adult-like) level of language production. Overall complexity scores are based on a simple count of the number of items on which parents checked the more complex option, permitting a range from o to 37 points (children who are still not combining words at all are given a default score of o on this scale). Critical to our purposes here, there are now several validation studies in both languages, showing that the grammatical complexity scale is highly correlated with free speech and/or sentence elicitation tasks in a laboratory setting (e.g. Dale, 1991; Devescovi, Caselli & Bonanni, 1996).

The 37-item complexity scales in English and Italian are similar in size, coverage, and sensitivity to contrasts that emerge in this age range, in each language. However, the two scales are certainly not translations, and their content is quite different. Because the Italian system of inflectional morphology is so rich, a separate section was developed for the Italian CDI to examine verb conjugation and noun declension paradigms (we will not consider that subscale in this cross-linguistic paper, because there is no comparable scale for English). Hence the Italian and English versions of the grammatical complexity scale both concentrate primarily on changes in morphosyntax that serve to increase sentence length and complexity. Table 2 lists some sample items from each scale, in English vs. Italian (for a complete listing, see Fenson *et al*., 1993, for English and Caselli & Casadio, 1995, for Italian). Clearly these scales favour those aspects of morphosyntax in which the two languages are most comparable, and are not sensitive to the larger set of morphological contrasts that Italian children have to acquire. We will return to this point later, in providing a qualitative look at the sentence types that American and Italian children produce at comparable levels of vocabulary development.

Finally, the grammar section of the CDI also asks parents to write out the three longest sentences that they can remember their child saying in the last couple of weeks (on the grounds that these would be sufficiently recent and

TABLE 2. *Sample items from the grammatical complexity checklists for English vs. Italian*

English

1.	Two shoe	22.	Where mommy go?
	Two shoes		Where did mommy go?
3.	Daddy car	26.	Don't read book
	Daddy's car		Don't want you read that book
5.	I make tower	28.	I want that
	I making tower		I want that one you got
12.	(talking about something that already happened)	35.	We made this
	Kitty go away		Me and Paul made this
	Kitty went away		
16.	You fix it?	37.	Baby crying
	Can you fix it?		Baby crying cuz she's sad

Italian (*with translation*)

1.	Scotta pappa	(It) burns, food
	Scotta la pappa	(It) burns, the food
2.	Scarpe mamma	Shoes mommy
	Scarpe di mamma	Shoes of mommy
		(possessive construction in Italian)
5.	Bimbo cade	Baby falls
	Bimbo cade per terra	Baby falls on ground
12.	Bimbo più	Baby more
	Bimbo non c'e più	Baby not any more
		(Italian expression for 'allgone')
16.	Via treno rosso nonno	Byebye train red grandpa
	Sono andato sul treno rosso col nonno	(I) went on the red train with grandpa
22.	Metto scarpe, via	(I) put shoes, byebye
	Mi metto le scarpe e vado via	(I) reflexive-put the shoes and (I) go byebye
26.	Bevo latte, nanna	(I) drink milk, nightnight
	Bevo il latte e dopo vado a nanna	(I) drink the milk and then (I) go nightnight
34.	Papà detto non mangia mella	Daddy (has) said not eat candy
	Papà ha detto che non si mangia la caramella	Daddy has said that not reflexive-eat the candy
37.	Lavo mani, sporche!	(I) wash hands, dirty!
	Mi lavo le mani, perche sono sporche!	(I) reflexive-wash the hands, because (they) are dirty!

striking events to have some validity even in recall mode). Some examples from this part of the grammar section will be reported later, for a small number of individual children. As we will see, these reported utterances are similar in complexity and structure to the sentence pairs in the complexity scale; more importantly, they are also similar to sentences that we and others have observed in independent free speech studies of children in this age range, in each language.

Data reduction

Before describing the calculation of proportion scores based on adult part-of-speech categories, we need to provide an important caveat, discussed in some detail in the Caselli *et al.* paper on younger children. That is, WE DO NOT ASSUME IN THE ANALYSES DESCRIBED BELOW THAT CHILDREN IN THIS AGE RANGE NECESSARILY REPRESENT OR RECOGNIZE CATEGORIES LIKE NOUN, VERB, OR GRAMMATICAL FUNCTION WORD. It has been known for some time that children in this age range sometimes use words in ways that deviate from adult patterns of usage for the very same word (Bates, 1976; Greenfield & Smith, 1976; Tomasello, 1992). For example, English-speaking adults may use the word 'hot!' as an adjective, to warn children about hot objects like a stove or light bulb; however, some children learning English go through a period where they use the word 'hot' as a label, to name light bulbs, candles, stoves and other potentially hot objects (Volterra, 1979; Volterra & Caselli, 1986). In view of this problem, what right do we have to analyse the child's reported vocabulary from the point of view of adult part-of-speech categories?

Our solution to this problem is to treat the adult categories as INDEPENDENT VARIABLES. That is, we treat these categories as a summary of the child's linguistic input, including similarities and differences between Italian and English in the phonological, semantic, morphological, and syntactic properties of nouns, verbs and grammatical function words. To the extent that children treat nouns differently from verbs, content words differently from function words, and so forth, we can assume that they have been affected by these differences in input. Hence developmental sequences in the acquisition of part-of-speech types may be taken to reflect changes in the child's ability to deal with these types. They do not necessarily reflect the emergence of explicit or implicit categories like noun or verb from the child's point of view. In the same vein, cross-linguistic differences in vocabulary composition (with vocabulary size held constant) can be taken to reflect the child's sensitivity to variations in the nature of the input language. They do not necessarily reflect cross-linguistic difference in the 'moment' (if there is such a moment) at which such categories emerge in the mind of the child.

The large category of common nouns was constructed by adding together the entries for each child in the following categories: animals, toys, food and drink, clothing, body parts, small household objects, furniture, and rooms. Following Bates *et al.* (1994) and Caselli *et al.* (1995), other possible nominals from the categories 'sound effects', 'places to go', 'names for people', 'time words', and 'games and routines' were excluded from this count in order to provide a conservative estimate of words that serve a clear naming function, refer to a class of nameable objects (as opposed to a single individual), and are minimally ambiguous with other form classes (e.g. many items in the 'places

to go' category behave like adverbials or locatives in the adult language). With 263 common nouns in each language, this class represents 38·7% of the 680 words on the English checklist, and 39·25% of the 670 words available in Italian. These percentages thus represent the checklist ceiling for common nouns in each language, i.e. the score for any child whose parents checked all the words on the list, and the average figure that we would expect if growth were randomly distributed across all classes.

The predicate category was constructed by adding together scores for verbs ('action words') and adjectives, yielding a total of 166 possible items, representing 24·4% of the English scale and 24·8% of the Italian scale. Potential predicative forms from the categories 'sound effects', 'places to go' and 'games and routines' were not included in this count, in order to avoid ambiguity and insure that the predicate category is a conservative estimate of words that serve as main verbs and noun modifiers in English and Italian. The verb category alone has 103 items in each language, so the checklist ceilings for the verb proportion scores are 15·14% in English and 15·37% in Italian.

The closed-class category (as described above) has 102 words in English and 92 words in Italian, yielding checklist ceilings of 15% and 13·7%, respectively (see Table 3).

Finally, we grouped together items from the three categories 'sound effects', 'names for people' and 'games and routines' into a single category that we will refer to as 'social words'. This category was not used by Bates *et al.* (1994), who excluded all three subsets from their compositional analyses, together with two other heterogeneous and ambiguous categories that we have also excluded from our analyses, i.e. 'places to go' and 'time words'. However, in their comparative study of the first stages of lexical development. Caselli *et al.* noted that the categories 'sound effects', 'names for people' and 'games and routines' contain many of the 'pure performatives' that predominate in the earliest stages of lexical development for many children. Furthermore, Caselli *et al.* noted some striking cross-language differences in this cluster of items, with higher ratios of social-word use among the Italian infants. Hence we decided to include the social-word category in our cross-linguistic analysis of the subsequent phase of development, from 1;6 to 2;6. There are 66 of these items on each list, comprising 9·7% of all English items and 9·85% of all Italian words.

In calculating these scores (and in the original design of the Italian and English scales), every effort was made to use comparable criteria for word classification in each language, consistent with our claim that adult part-of-speech categories should be treated as independent variables in the analysis of child language. Nevertheless, because the adult categories themselves are idealizations, we were forced to make some arbitrary (and arguable) classifications in many cases, especially for routines, sound effects and other

85

marginal categories (i.e. items that one cannot simply look up in a dictionary). There are a few items that ended up in different categories for English versus Italian, even though they have a similar function in each language. For example, the word 'uh-oh' is classified under sound effects in English, whereas its nearest equivalent 'bum' is classified as a routine in Italian. All of these classifications are probably best viewed as approximations, which means that we should be very sceptical of cross-language differences that could turn on one or two items within a given category.

RESULTS AND DISCUSSION

We will begin with a global comparison of age-related changes in expressive vocabulary in the two groups. Then we will examine similarities and differences between English and Italian in the composition of vocabulary, with children grouped according to vocabulary size rather than age. This is the section that will provide our most important test of the predicated verb advantage in Italian. It also includes information about cross-linguistic variation in function word growth.

Having completed our survey of the pace and composition of vocabulary development in this age range, we will look at changes in sentence complexity in English and Italian, as a function of age and vocabulary level. Then we will take a qualitative look at the acquisition of individual closed-class words in each language, pointing out specific examples where content may vary across languages. Finally, examples are provided of the three longest utterances reported for a subsample of children in the two language groups, matched for age, sex and approximate vocabulary size. This last analysis will also help to illustrate the range of variation in content that can be observed, even though the overall pace and shape of lexical and grammatical development is quite similar in these two languages.

Cross-linguistic analyses of vocabulary composition

To determine whether the two groups differ in rate of vocabulary development from 1;6 to 2;6, we carried out a (2)language × (13)age between-group analysis of variance. Results are illustrated in Figure 1.

Not surprisingly, there was a large and reliable main effect of age ($F(12, 1325) = 65 \cdot 11$, $p < 0 \cdot 0001$). There was also a significant main effect of language ($F(1, 1325) = 31 \cdot 02$, $p < 0 \cdot 0001$). The interaction was not significant ($F(12, 1325 = 1 \cdot 21$, n.s.). As can be seen from Figure 1, Italian children lag behind their English-speaking counterparts at almost every age level. For example, the average vocabulary score at 1;6 is 69 for Italians vs. 114 for English. The corresponding figures at 2;6 for 446 words for Italian, and 534 words for English. We have no ready explanation for this apparent 'English advantage', which also appeared in the Caselli *et al*. study of younger children. The Italian checklist in the present study is 10 words shorter than
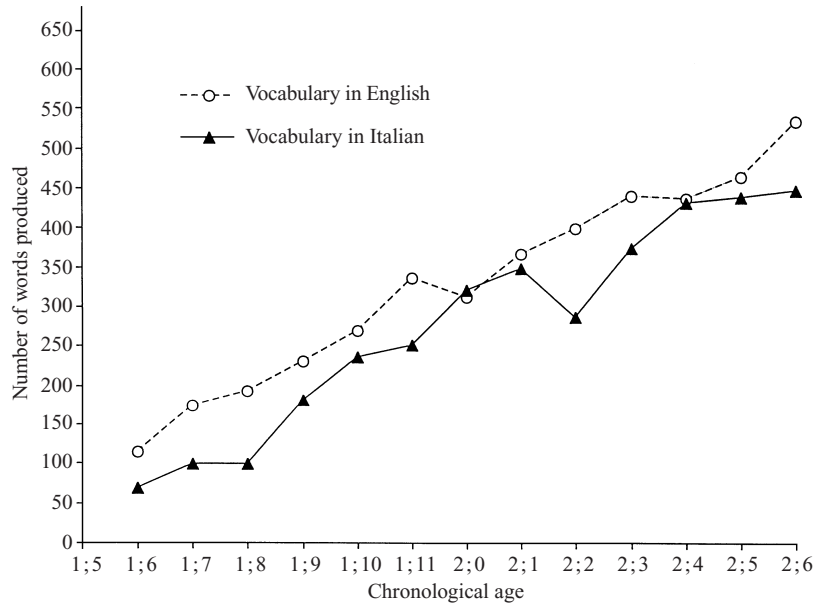
Fig. 1. Vocabulary size as a function of age in English and Italian children.

its English counterpart, but this difference is due entirely to items within the closed class, and (as we shall see below) there is no evidence that the English advantage comes from that category. Whatever its cause, a quantitative difference between groups in rate of development could obscure the more qualitative variations that are the focus of this study (e.g. variations in noun vs. verb growth; variations in the relationship between vocabulary size and grammar).

To control for this global difference in rate of vocabulary growth, we are faced with two alternatives: (1) match individual Italian children for age, gender and vocabulary size with individual children from the larger English sample, or (2) conduct all analyses with children grouped by vocabulary size rather than age. The first alternative has two disadvantages: it decreases statistical power, and (even with samples this large) is difficult to achieve on a case-by-case basis. The second alternative has two advantages: it maximizes statistical power (i.e. no information is thrown away), and facilitates comparison with our other published studies using this procedure (Bates *et al.*, 1994; Caselli *et al.*, 1995). Because both alternatives achieve the same goal, we have opted for the second in all analyses presented below.

For analyses of vocabulary composition, children were divided into eight groups based on their total vocabulary size (following Bates *et al.*, 1994): (1) 1–50 words, (2) 51–100 words, (3) 101–200 words, (4) 201–300 words, (5)

87

301–400 words, (6) 401–500 words, (7) 501–600 words, and (8) more than 600 words. A breakdown of the sample by language and vocabulary level is provided in Table 3.

TABLE 3. *Breakdown of the two samples by vocabulary level*

	English		Italian	
Vocabulary level	Number	Sample (%)	Number	Sample (%)
1–50 words	88	8·8	62	16·1
50–100 words	96	9·6	48	12·4
101–200 words	154	15·4	52	13·5
201–300 words	128	12·8	65	16·8
301–400 words	152	15·2	46	11·9
401–500 words	154	15·4	49	12·7
501–600 words	145	14·5	48	12·4
> 600 words	84	8·4	16	4·1
Totals	1001		386	

To confirm that this division did indeed equate the two language groups for vocabulary size, we conducted a (2) language by (8) vocabulary level between-group analysis of variance using total word production as the dependent variable. There was a huge main effect of vocabulary level, which is of course entirely circular, since vocabulary level was defined by the dependent variable ($F(7, 1324) = 10071$, $p < 0.0001$). Most important for our purposes, there was no significant main effect of language ($F(7, 1324) = 1.27$, n.s.), and no interaction between language and vocabulary level ($F(7, 1324) = 1.42$, n.s.). This means that we were successful in balancing the two groups for vocabulary level, and can proceed to a description of language by level effects on vocabulary composition.

Figure 2*a* illustrates changes as a function of vocabulary level in English, for percent common nouns, percent predicates, percent closed-class words, and percent social terms. The first three proportion scores have already been reported for this English sample in Bates *et al.* (1994); the social-word scores are an addition that we developed in our earlier cross-linguistic study (Caselli *et al.*, 1995), to capture some small but compelling differences between the two language groups. The flat dotted lines in this figure (and those that follow) indicate the checklist ceiling for each category, i.e. absolute proportion of words from that category on the checklist as a whole. For example, common nouns comprise just under 40% of all words on the checklist, verbs and adjectives around 24%, function words between 14 and 15%, and social words just under 10% (recall that time words and 'places to go' are excluded from these counts, which explains why the proportions do not add up to 100%). This disproportional representation reflects both the unequal distri-
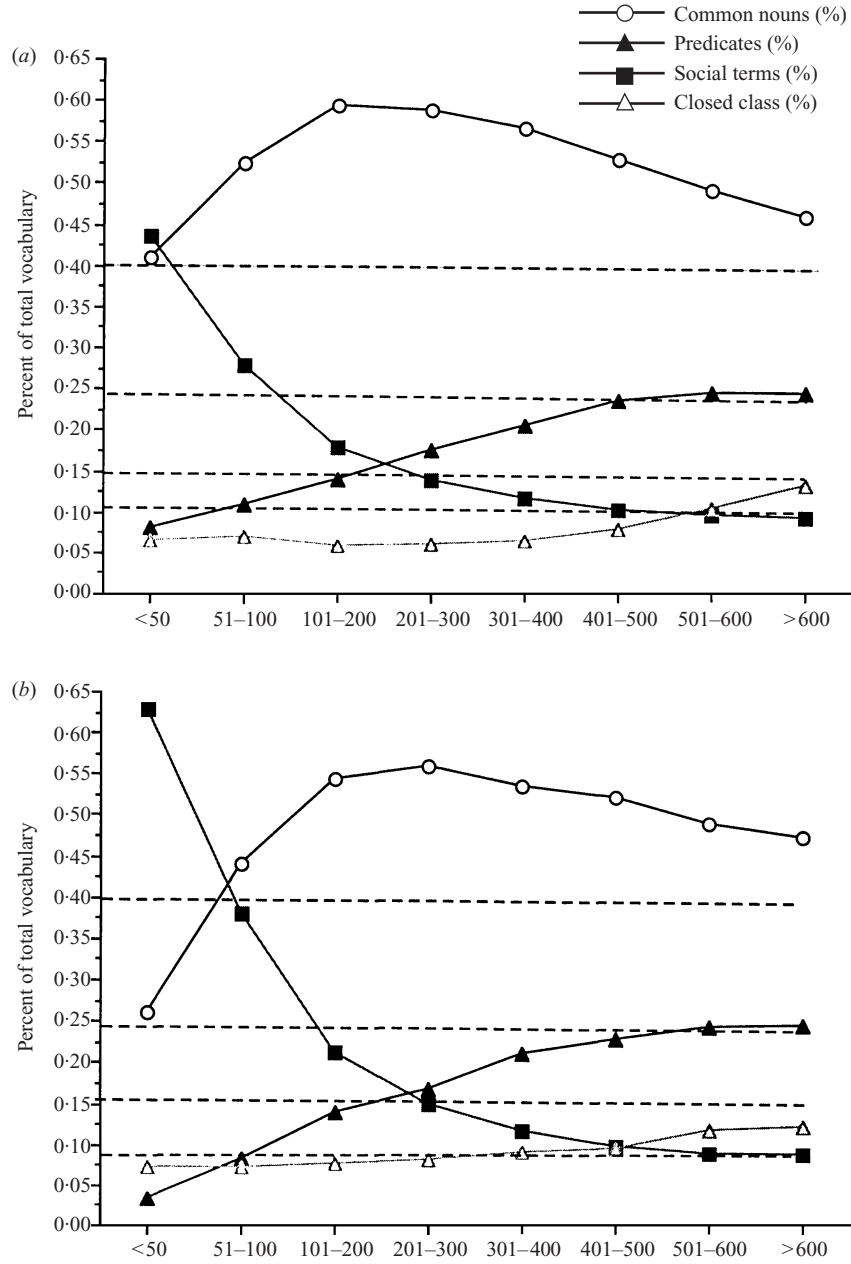
Fig. 2(a). Vocabulary composition for American children from 1;6 to 2;6. (b) Vocabulary composition for Italian children from 1;6 to 2;6.

bution of words in the adult language (e.g. there are many more nouns than verbs, and many more verbs than function words in both English and Italian), and the unequal proportions that one finds in all diary and free speech studies of early development. In order to assess developmental changes in proportion scores, we need to assess development against this background. If development proceeded evenly across all word classes, with words added in accordance with their representation on the checklist as a whole, then the developmental functions would be flat (hovering around 40 % for common nouns, 24 % for verbs and adjectives, and so forth). Instead, as can be seen from Figure 2*a*, common nouns show a marked increase between 1 and 200 words, with a gradual drop after that point. Predicates show continuous growth across this period of development, levelling off near the checklist ceiling for children with vocabularies of 400 words or more. Closed-class words show no growth at all in English up to 400 words, and then start to rise to their checklist ceiling after that point. Finally, social words start out as the largest category of all for children with vocabularies under 50 words (slightly above the high values for common nouns), but this category shows a precipitous nonlinear drop after that point.

Figure 2*b* illustrates the corresponding data for children acquiring Italian. A quick comparison of Figures 2*a* and 2*b* indicates that the four variables show similar overall patterns of growth across vocabulary levels in English and Italian. The most noteworthy similarities are the preponderance of common nouns, the slow growth of predicates (which do not outnumber nouns at any point), the rarity of closed-class words in early vocabularies, and the sharp nonlinear drop in social-word proportion scores after the earliest level.

Despite these global similarities, Figures 2*a*, *b* suggest that there may be subtle differences in the shape of these functions. To explore these differences, five separate (2) language × 8 vocabulary level analyses of variance were conducted, on proportion scores for common nouns, predicates (verbs and adjectives combined), closed-class words, and social words, respectively. We also conducted a separate analysis on verbs alone, to provide the clearest possible test of the prediction that verbs will develop earlier in Italian. When significant language by level interactions appear, simple-effects analyses (one-way analyses of variance by language) were carried out at each vocabulary level, to uncover the locus of the interaction. In all instances, the p-value for these *post hoc* analyses was $p < 0.05$.

The common-noun analysis yielded significant main effects of vocabulary level ($F(7, 1324) = 147.34$), and language ($F(1, 1324) = 92.04$, $P < 0.0001$), and a significant language by vocabulary level interaction ($F(7, 1324) = 12.88$, $p < 0.0001$). This interaction is illustrated in Figure 3, which shows that proportional growth in the common-noun category has a similar shape in English and Italian; however, the function is somewhat higher for
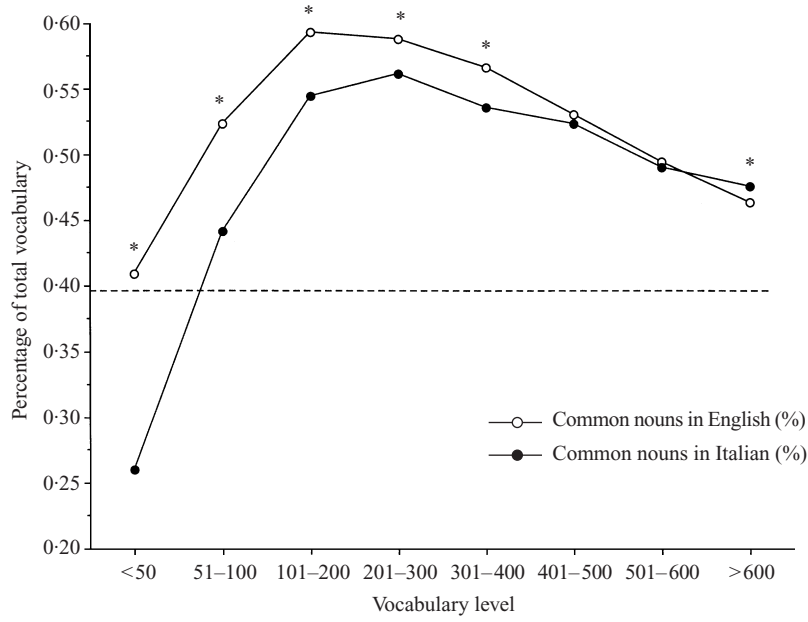
Fig. 3. Common nouns as a proportion of total vocabulary size (dotted line = checklist ceiling).
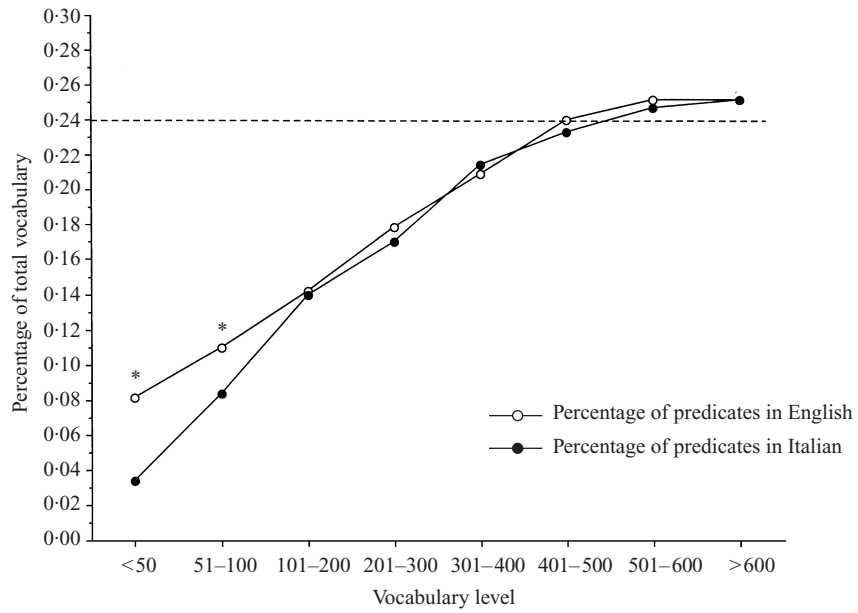

Fig. 4. Predicates as a function of total vocabulary size (dotted line = checklist ceiling).

91

Americans, especially among children with fewer than 200 words. Simple-effects analyses showed that the two groups differ at all points except the two levels between 400 and 600 words. The advantage goes to the American sample in every case except for the final level (over 600 words), where there is a very small but reliable difference favouring the Italians.

The analysis of variance over predicates also yielded significant main effects of level ($F(7, 1324) = 354.83$, $p > 0.0001$) and language ($F(1, 1324) = 14.81$, $p > 0.0001$), plus a significant interaction ($F(7, 1324) = 3.90$, $p > 0.0001$). This interaction is plotted in Figure 4. Simple-effects analyses show that predicate scores are significantly larger for the American group at the first two vocabulary levels, contrary to predictions of a 'verb advantage' in Italian, but similar to findings reported by Caselli *et al.* for younger children. There are no significant language differences after that point.

To clarify whether this does indeed mean that there is no verb advantage in Italian, we repeated the analysis of variance for verb proportion scores only. This analysis also yielded significant main effects of level ($F(7, 1324) = 264.27$, $p < 0.0001$) and language ($F(1, 1324) = 8.12$, $p < 0.004$), plus a small but reliable interaction ($F(7, 1324, = 2.21$, $p < 0.04$), illustrated in Figure 5.
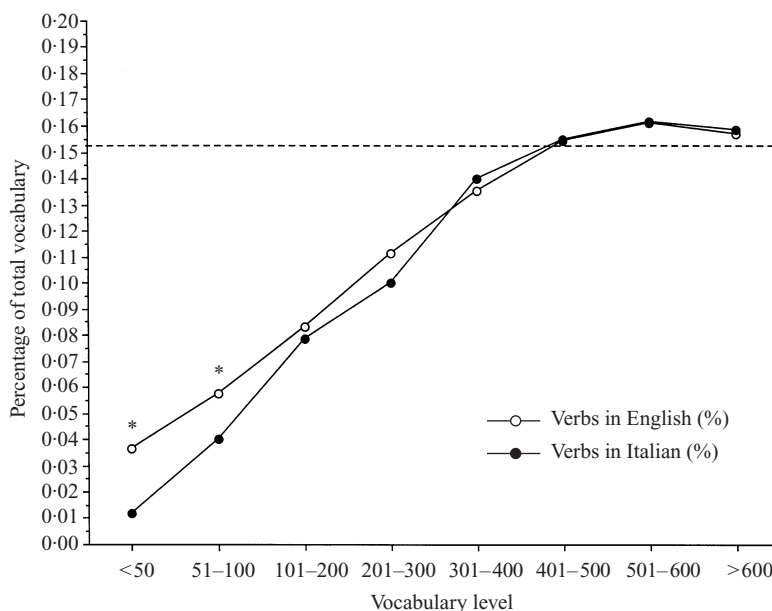

Fig. 5. Verbs as a function of total vocabulary size (dotted line = checklist ceiling).

This figure shows that the verb advantage does favour English-learning children, contrary to predictions, although the difference is only reliable at the first two vocabulary levels.

So far, it looks as though the American children are ahead on all the major categories. Because these are proportion scores, we know that the Italians must be making up the difference somewhere else. An analysis of variance on the social-word category yielded significant main effects for level ($F(7, 1324)$ $= 788.77$, $p < 0.0001$) and language ($F(1, 1324) = 119.99$, $p < 0.0001$), as well as a language by level interaction ($F(7, 1324) = 44.71$, $p < 0.0001$). The interaction is plotted in Figure 6, which shows that the Italians have a clear
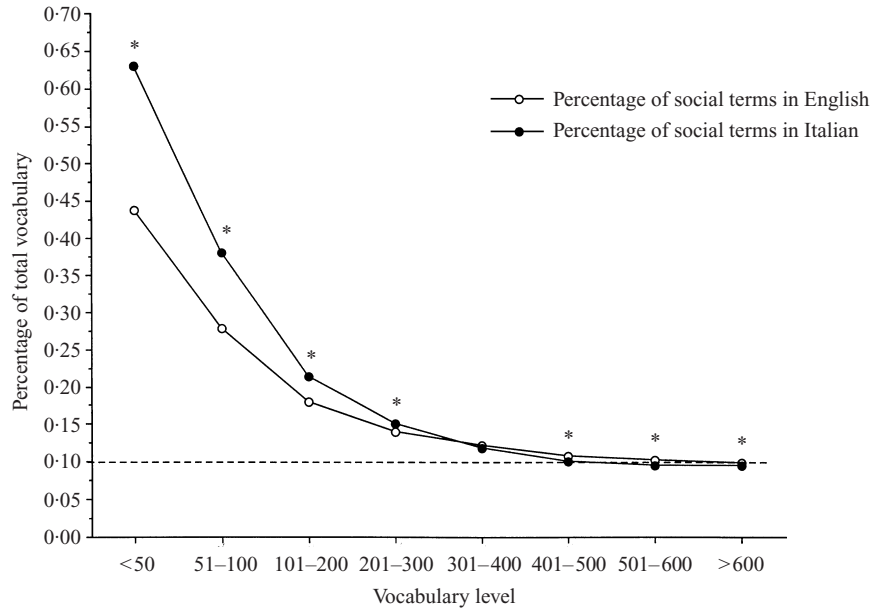

Fig. 6. Social terms as a proportion of total vocabulary size (dotted line = checklist ceiling).

advantage in the social-word category in the early phases of lexical de-velopment. The contribution of social words to total vocabulary undergoes a sharp nonlinear drop for both languages, as we have already noted. How-ever, *post hoc* analyses showed that the Italians maintained their advantage in this category at seven of the eight vocabulary levels. This finding replicates our earlier report for younger children. In that study, we suggested that the greater representation of social words in Italian reflects cultural differences, including the tendency for Italian families to live in the same cities with an extended family, a fact that gives Italian children more relatives to be named and more relatives to elicit routines, sound effects and other language games on their frequent visits. Because our analyses are all based on proportion

93

scores, the relatively greater number of social words in the Italian sample occurs at the cost of relatively smaller scores for content words in this language.

Finally, the analysis of variance for closed-class proportion scores yielded significant main effects of level ($F(7, 1324) = 54.21$, $p < 0.0001$) and language ($F(1, 1324) = 31.56$, $p < 0.0001$). The interaction was also reliable, although it was relatively small ($F(7, 1324) = 2.11$, $p < 0.04$). The shape of this interaction can be seen in Figure 7, which shows that Italian children are
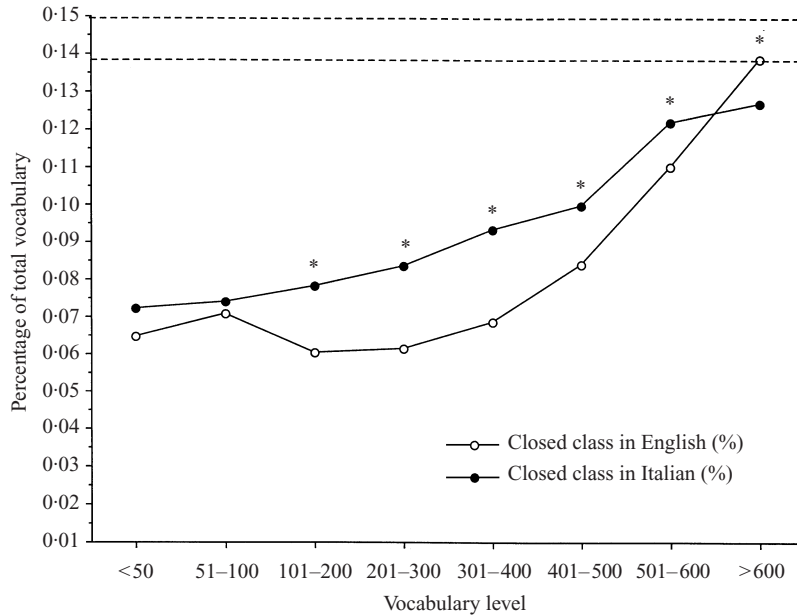

Fig. 7. Closed-class words as a proportion of total vocabulary size (dotted lines = checklist ceiling for English (15 %) and Italian (13.7 %).

slightly ahead of their American counterparts at every level, except for the final level when vocabularies exceed 600 words. *Post hoc* tests showed that this difference is not reliable for children with vocabularies under 100 words, but it is reliable at all levels after that point.

In addition to these differences in the magnitude of closed-class proportion scores. Figure 7 shows that the overall shape of the growth function for closed-class words is also somewhat different in these two languages. As Bates *et al*. (1994) have already shown, the English closed-class function is nonlinear; there is no detectable relationship between these proportion scores and total vocabulary size until the point at which total vocabulary exceeds 400 words (see Bates *et al*., for a detailed discussion of this point). By

94

contrast, the Italian function approaches linearity, with gradual increases across the entire period in the proportional contribution of closed-class items to total vocabulary. To explore these apparent differences in shape, we conducted separate one-way analyses of variances on closed-class proportion scores as a function of vocabulary level within each of the language groups. In the Italian analysis, the weighted linear function was highly reliable ($F = 45 \cdot 16$, $p < 0 \cdot 0001$), and there was no significant deviation from linearity ($F = 0 \cdot 84$, n.s.). In the English analysis, there was also a reliable weighted linear component ($F = 308 \cdot 44$, $p < 0 \cdot 0001$), but in this case the component assessing deviation from linearity did reach significance ($F = 25 \cdot 81$, $p < 0 \cdot 0001$). A similar result is obtained with correlational analyses. Within the American group, there is no correlation between closed-class proportion scores and vocabulary size in children with fewer than 400 words, whether we look at vocabulary level ($r = +0 \cdot 01$, n.s.) or total vocabulary in words ($r = +0 \cdot 02$, n.s.). By contrast, the corresponding correlations for children under 400 words did reach significance in our Italian sample (with vocabulary level, $r = +0 \cdot 13$, $p < 0 \cdot 04$; with total vocabulary in words, $r = +0 \cdot 14$, $p < 0 \cdot 03$). These analyses confirm the apparent differences in the shape of the two functions.

To summarize results for vocabulary composition, we find no evidence for a verb advantage in Italian, replicating and extending our previous cross-linguistic study of younger children. The overall shape of development for both nouns and verbs is quite similar in these two languages. There are small differences favouring the Americans in both these categories, but these can be attributed to the fact that Italian children have a correspondingly larger repertoire of social words. We do find a small but consistent advantage for Italian children in the proportional development of closed-class words, complemented by small but reliable differences in the shape of change within this category. Later on, we will present a qualitative breakdown of the acquisition of specific function words, shedding some light on the different growth patterns for function words in English and Italian.

Cross-linguistic analysis of grammar in relation to vocabulary

We begin with a (2) language by (13) age between-group analysis of variance of scores from the grammatical complexity scale (with a possible range from 0–37 in each language). Results included significant main effects of age ($F(1, 1325) = 58 \cdot 26$, $p < 0 \cdot 0001$), and language ($F(12, 1325)$ $10 \cdot 17$, $p < 0 \cdot 001$), as well as a significant age-by-language interaction ($F(12, 1325) = 2 \cdot 53$. $p < 0 \cdot 003$). The interaction is illustrated in Figure 8, which indicates an advantage in grammatical complexity for American children at 10 of the 13 data points. Notice also that the curve for English tends to be smooth and monotonic, while the Italian data are far less consistent, with a number of surprising ups and downs. We believe that this inconsistency reflects two
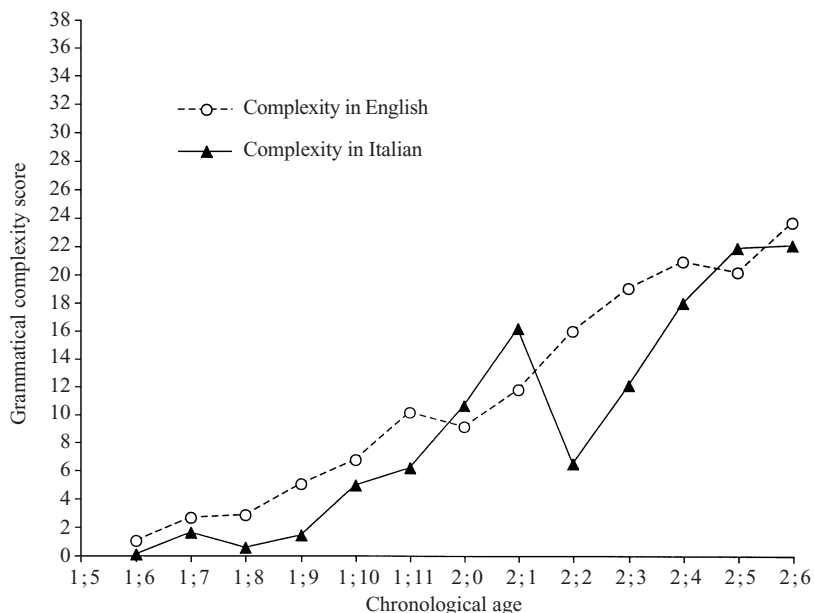
Fig. 8. Grammatical complexity as a function of age.

separate facts: the greater stability of the American data (due to larger sample size), and the fact that age alone is not a particularly good predictor of grammatical development within this age range (Fenson *et al.*, 1994; Bates & Goodman, 1997). The language difference is now quite familiar to us, since it appeared in the analysis of age by total vocabulary size described above, and in our earlier cross-linguistic study of word production (Caselli *et al.*, 1995). The important question for our purposes here is whether this cross-language difference will persist when the two groups are equated for total vocabulary.

Towards that end, we conducted a (2) language by (8) vocabulary level between-group analysis of variance on the same grammatical complexity scores. This analysis revealed a very large and reliable main effect of vocabulary level ($F(7, 1324) = 44830$, $p < 0.0001$). There was no main effect of language ($F(1, 1324) = 1.28$, n.s.), and no significant interaction ($F(7, 1324) = 1.42$, n.s.). Although the interaction is not reliable, we have plotted grammatical complexity as a function of vocabulary level for both languages in Figure 9, to facilitate comparison. This figure shows that the nonlinear function linking grammatical development with vocabulary size is remarkably similar in English and Italian, despite clear differences between these languages along grammatical dimensions that include richness of morphological marking and degree of permissible word order variation. This is a compelling cross-linguistic replication and confirmation of previous
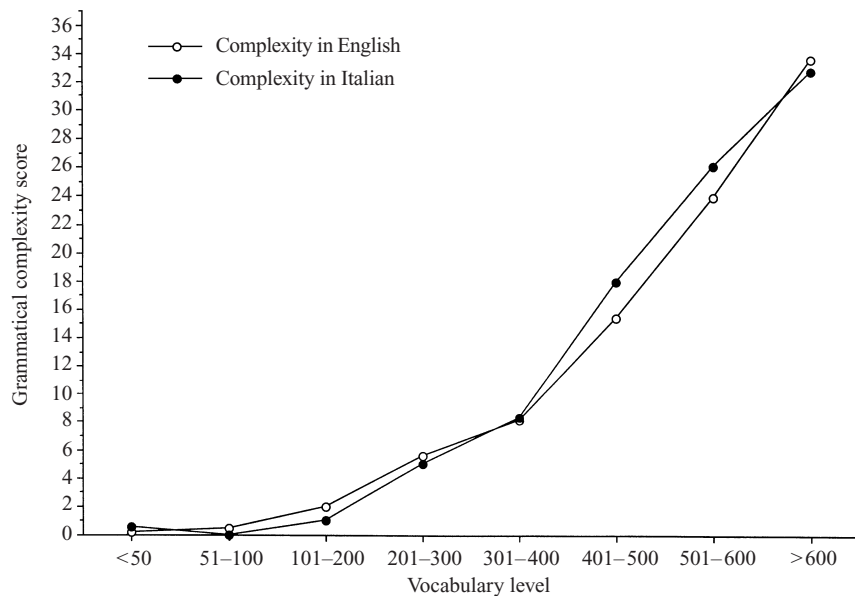
96

Fig. 9. Grammatical complexity as a function of vocabulary size.

reports for English, suggesting that there is a powerful and perhaps causal link between lexical development and the emergence of grammar in this crucial period of development (Fenson *et al.*, 1994; Marchman & Bates, 1994; Dale, 1996; Bates & Goodman, 1997).

Of course it is important to keep in mind that we constructed these complexity scales to facilitate comparison over languages. Both scales contain exactly 37 pairs of sentences, and both are constructed to pick up those changes that are known to occur between 1;6 and 2;6, based on many naturalistic studies of development within each language. These are both Indo-European languages, and the specific structural dimensions that are contrasted in the 37 sentence pairs tend to be those that occur across the whole language family (e.g. presence/absence of plural and past tense marking; simple sentences vs. sentences with conjoined or embedded elements). However, the items on the two respective scales are not translations, and each scale includes many contrasts that are not available in the other. Hence the close resemblance between the two grammar-on-vocabulary functions in Figure 9 is not at all trivial (see Table 2).

At the same time, we must acknowledge that these two growth functions could be driven apart if the two scales had been designed explicitly to maximize the most important differences between English and Italian. For example, we know *a priori* that Italian children will have to master a larger array of morphological contrasts than their English-speaking counterparts

97

(for a detailed discussions, see Bates, 1976; Leonard, Bortolini, Caselli, McGregor & Sabbadini, 1992; Pizzuto & Caselli, 1992; Cresti & Moneglia, 1993; Devescovi & Pizzuto, 1995; Devescovi *et al.*, 1996). If we had a scale that could tap directly into these differences in morphological complexity, interesting cross-language differences in the shape of the relationship between grammar and vocabulary size would undoubtedly emerge. We will return to this point later.

Because our analyses of vocabulary composition suggested that there are subtle differences between English and Italian in the growth curves for function words (relative to vocabulary size), we decided to examine the sequence of acquisition for specific function words within each major category, for each language. Table 4 presents a full listing of the grammatical function words contained on each of the word checklists, organized by language (English on the left, Italian on the right) and type (pronouns and pronominal determiners, question words, prepositions, quantifiers and articles, connecting words and auxiliary verbs). Within each category, words are listed in their order of acquisition. Following Fenson *et al.* (1994), Caselli *et al.* (1995), and Caselli & Casadio (1995), age of acquisition is operationalized as the percentage of all children within each language who are reported to produce that item. As Fenson *et al.* have shown, this simple statistic correlates highly with more complicated month-by-month estimates (e.g. the age at which 50 % of the sample is reported to produce a given word, with adjustments for those words that are still not mastered by 50 % of the sample by the end of the study at 2;6). Because minute differences in any of these scores could be affected by variations in sample size and method (e.g. the fact that there are 10 more closed-class words on the English list), we will not attempt a statistical analysis of these order-of-acquisition data. The reader is invited to examine Table 4 for details; we will restrict ourselves here to a qualitative summary of results within and across function word categories.

The most important finding in Table 4 is the high degree of similarity in order of acquisition of function words in English and Italian, even though the content of the two lists is not identical (and comparisons are simply not possible in some cases, e.g. the fact that Italian has multiple reflexive and clitic pronouns that have no counterpart in English). For example, singulars tend to come in earlier than plurals in every relevant class (e.g. pronouns, including pronominal determiners; auxiliaries). The pronominal determiner 'Mine!' (Italian 'mio') is the first item in the pronoun class in both languages, and 'more' (Italian 'ancora') is the first quantifier, facts that may reflect universal social and material concerns of one-year-olds. Within the pronoun class, person marking follows the same sequence in both languages, with first person < second person < third person (though Italians show a marked delay in informal second person plural forms like 'voi', 'vi' and

TABLE 4. *Percentage of children reported to produce specific function words in English and Italian in English and Italian*

Pronouns (including pronominal determiners)		Pronomi	
Mine	78·7	Mio	82·38
Me	67·1	Io	66·58
That	55·2	Tuo/a	56·74
You	54·0	Quello/a	53·63
I	53·9	Questo/a	52·85
My	50·1	Tu	48·96
This	47·9	Me/Mi	38·34
It	42·7	Si	35·75
He	29·8	Te/Ti	27·20
These	26·5	Suo/a	26·68
She	24·8	Lo/a	24·09
Your	23·5	Che	22·80
Her	22·1	Lui	22·54
We	21·0	Lei	18·91
Him	20·6	Noi	18·65
Myself	20·0	Li/e	18·13
Those	18·5	Nostro/a	13·73
His	18·1	Ci	12·95
Them	16·9	Gli	9·59
They	15·7	Loro	8·29
Hers	15·5	Voi	5·70
Our	14·1	Vi	4·92
Us	11·2	Vostro/a	2·85
Their	11·0		
Yourself	8·3		
Question words		**Interrogativi**	
What	54·8	Che/Che cosa?	37·82
Where	44·3	Chi	35·75
Why	35·7	Dove?	35·49
Who	30·6	Perché?	31·61
How	21·9	Come?	17·62
When	14·2	Quando?	15·03
Which	8·5	Quale?	13·73
Prepositions and locations		**Preposizioni**	
Down	79·4	Ecco	59·33
Up	76·2	Qui/Qua	59·07
Off	70·4	Giù	57·77
Out	69·4	Lì/Là	56·48
On	68·8	Fuori	46·11
Inside/In	58·5	Sotto	45·34
Here	50·7	Su	43·52
There	45·2	A	43·26
Back	42·5	Di	42·23
Away	37·9	Dentro	40·16
Under	36·4	Sopra	39·64
Over	35·2	Da	33·68

TABLE 4. (*cont.*)

Propositions and locations		Preposizioni	
To	34·9	Con	28·76
With	31·1	Lontano	26·94
At	29·8	Vicino	24·97
For	27·9	In	23·83
On top of	25·8	Dietro	23·32
By	24·7	Per	22·28
Around	23·2	Davanti	21·24
Behind	22·6	Fra/Tra	5·96
Next to	16·0		
Of	13·7		
Into	12·6		
Above	9·6		
About	9·6		
Beside	6·9		

Quantifiers and articles		Articoli e quantificatori	
More	75·3	Ancora	58·81
Too	45·8	Tanto	48·96
Some	44·2	Tutto	47·41
All	41·4	Poco	46·89
A	34·5	La	40·67
The	32·5	Altro/Un altro	39·90
Not	29·9	Un/Uno//Una	37·31
Other	29·7	Niente	34·20
Another	29·6	Il	31·87
Any	23·9	Un po'	31·61
A lot	23·6	Anche/Pure	28·24
None	19·4	I	27·20
Same	17·9	Nessuno	23·83
Much	16·9	Molto	22·28
Every	12·7	Lo	21·50
An	8·7	Le	21·24
Each	7·2	Di più	18·91
		Troppo	16·84
		Del/Della	15·28
		Dei/Delle	10·36
		Gli	9·84

Connecting words		Congiunzioni	
And	34·6	Perché	41·71
Because	20·9	E	30·57
So	15·6	Così	27·20
But	12·0	Ma	20·21
Then	11·0	Quindi/Allora	19·43
If	8·1	Se	10·62

Helping verbs		Ausiliari	
Do	54·3	Voglio	45·08
Wanna	49·1	È	39·12
Don't	45·4	Ho	31·35
Lemme	41·2	Sono	31·09
Gonna	33·2	Sei	26·17
Can	32·2	Posso	24·87
Did	29·9	Vuoi	24·87

TABLE 4. (*cont.*)

Helping verbs		Ausiliari	
Is	28·8	Ha	23·83
Need to	28·8	Vuole	21·24
Try to	26·0	Hai	20·73
Am	25·0	Devo	16·58
Have to	23·8	Puoi	10·62
Are	20·4	Devi	10·62
Be	19·4	Può	10·36
Gotta	18·5	Deve	9·07
Will	18·2		
Does	17·0		
Was	15·4		
Could	11·6		
Were	9·0		
Would	8·6		

'vostro'), and subject forms generally appear before their object counterparts (e.g. 'we', 'he' and 'she' precede 'us', 'her' and 'him', respectively). Question words appear in roughly the same order in both languages (What < where < why < how < when < which), although 'where' seems to be earlier in English while 'chi' ('who') is earlier in Italian. Connecting words follow the same sequence (['and', 'because'] < 'so' < 'but' < 'then' < 'if'), although 'and' precedes 'because' in English while the opposite order occurs in Italian. Prepositions and locatives show a number of parallels here that have also been reported in free speech and experimental studies (Johnston, 1985): words that express direction or location of a single element emerge first (e.g. 'down', 'up', 'off', 'out', 'here' and 'there'), followed by locatives that mark a simple relationship of one entity to its base ('on', 'inside', 'under', 'over'), while the locatives that appear last are those that express a relationship between two entities and/or a relationship that requires assumptions about the orientation of the array relative to the speaker and listener (e.g. 'next to', 'beside', 'behind'). Comparisons are harder to make within the categories of quantifiers and articles, and modals and auxiliaries, but there are similarities here as well. For example, the order 'want' < 'can' < ['have to', 'gotta'] in English corresponds to the acquisition of the first conjugated form of each modal in Italian ('voglio' or 'I want' < 'posso' or 'I can' < 'devo' or 'I must').

The few differences that remain can be explained by structural, statistical and/or pragmatic differences between the two languages, superimposed on universals of cognitive development and infant social life. For example, the English subject pronoun 'I' is the fifth pronominal form acquired, reported for 50·1 % of the sample. The corresponding Italian subject pronoun 'io' is the second pronominal form acquired, reported for 66·58 % of the sample. Note that the pronoun 'I' is far more frequent than the Italian 'io', due to

the prevalence of subject omission in Italian. Because subject pronouns can be omitted in Italian, 'io' is only used for contrastive purposes; hence, when it is used, it is generally high in pragmatic and acoustic salience. Although this modest cross-linguistic difference will require a more careful test under comparable laboratory conditions, it suggests that salience may be more important than sheer frequency during the first stages of grammatical development.

We would also like to underscore another similarity in Table 4 that transcends specific function word categories: Acquisition appears to be a gradual process that extends across the period from 1;6 to 2;6, with individual items acquired at different points in time depending on their frequency, regularity, salience and utility to the child. There is no evidence here for a single 'moment' when articles, pronouns, prepositions or auxiliary verbs come in together as a block. In this respect, and in the order in which specific items emerge within each category, our results are largely (though not perfectly) compatible with previous free speech studies of grammatical development in English (e.g. Brown, 1973) and Italian (e.g. Pizzuto & Caselli, 1992; Devescovi & Pizzuto, 1995).

This brings us at last to the transition from words to sentences. Although the MacArthur questionnaires cannot provide detailed evidence on sentence complexity in this age range, an indirect estimate of cross-language differences in morphological complexity can be obtained by examining that part of the questionnaire on which parents were asked to list the three longest utterances that they remember hearing their child say in the last few weeks. A detailed qualitative and quantitative analysis of the three longest utterances reported by more than 1300 parents would take us far afield (and is, in fact, the subject of work in progress). However, we think it would be appropriate to contrast the similarities in grammatical development illustrated in Figure 9 and Table 3 with just a few examples of the kinds of sentences that Italian and American parents report for children at comparable levels of vocabulary development. As we shall see, this kind of close, qualitative comparison reveals differences that are not captured by the measures that we have used so far.

To conduct these comparisons, we randomly selected 20 cases from the Italian data base, five boys and five girls at 2;0 and five boys and five girls at 2;6. For each of those cases, we attempted to find a match in age and vocabulary size from the American sample, in order to compare the three longest utterances reported by parents within each language when vocabulary size is held constant. For our purposes here, we offer a few examples from different points along the development continuum. Starting with our least advanced talkers in each subgroup, consider the similarities and differences that are observed for an Italian boy at 2;0 with 201 words and an American boy at 2;0 with 203 words. The three sentences reported for our Italian case

are the following (subscripts indicate verb tense, mood and aspect, as well as morphological contrasts that trigger agreement):

(1) Scotta pappa, non vojo.
Burns$_{3rd\,sing.}$ food$_{3rd\,sing.}$, no want$_{1st\,sing.}$.
(The food is not, I don't want it)

(2) Metti giacca, esco io e Dede.
Put$_{2nd\,sing.\,imperative}$jacket, go-out$_{1st\,sing.}$ I and Dede.
(Put on your jacket, Dede and I are going out.)

(3) Lavo mani, sporche, apri acqua.
Wash$_{1st\,p.\,sing.}$ hands$_{fem.pl.}$, dirty$_{fem.pl.}$, turn-on$_{2nd.sing.}$ water.
(I am washing my hands, dirty, turn on the water.)

It is clear that this boy has already begun to master some of the markings that are required in Italian for person (on verbs), number (on verbs, nouns and adjectives) and gender (on nouns and adjectives), although articles and clitic pronouns are missing in obligatory contexts. Contrast this with the three longest utterances reported for the corresponding English-language case, a child of the same age, gender and vocabulary size:

(4) Mommy, go outside.
(5) Mommy, juice please.
(6) Shower wet.

The English examples are shorter overall (i.e. fewer content and function words), and they also contain no evidence for morphological productivity. This does not mean, of course, that morphological marking is completely absent for all English-learning children in this range of development. Consider, for example, the three longest utterances reported for an American girl at 2;0 with 210 words:

(7) Daddy go to work.
(8) I want pancakes.
(9) Amy pushed me.

These sample sentences are still quite short, but they provide evidence of at least some move in the direction of grammatical morphology, in the provision of a past tense marker on the verb 'pushed' and a plural marker on the noun 'pancakes' (but note the missing third person singular marker on the verb 'go').

Moving ahead in our small subsample, the following three sentences are reported for an Italian girl at 2;6 with 572 words:

(1) Mamma non voglio andare all'asilo, ma a scuola con la Costanza.
Mommy, not want$_{1st.sing.}$ to-go to the $_{masc.sing.}$ preschool$_{masc.sing.}$, but to school with the$_{fem.sing.}$ Costanza$_{fem.sing.}$
(Mommy, I don't want to go to preschool, but to school with Costanza.)

103

(11) Mama, perché dondola quel signore, è malato?
Mommy, why rocks$_{3rd.sing.}$ that$_{3rd.sing.masc.}$ man$_{3rd.sing.masc.}$, is$_{3rd.sing.}$
sick$_{sing.masc.}$?
(Mommy, why is that man rocking, is he sick?)

(12) I denti ci si lavano dopo mangiato, vero, Mamma? Prima no.
The$_{masc.pl.}$ teeth$_{masc.pl.}$ (to-us) ($_{3rd.sing.reflexive}$) are-wash$_{3rd pl.}$ after
eaten$_{particle,}$ true, Mommy? Before no.
(Teeth get washed after eating, right Mommy? Not before.)

These examples illustrate several features that distinguish Italian from
English, including the prevalence of subject omission, the use of clitic
pronouns, and the rich morphological marking that is required even for
relatively simple sentences. This child is now providing most of the required
morphological contrasts, at least for the sentences reported here (it is
important to remind the reader that these sentences represent the parents'
memory of their child's very best efforts). As a result, these Italian examples
contrast sharply with the three longest utterances reported for an American
boy at 2;6 with 581 words:

(13) Daddy, can you help me find it please?
(14) I wanna help wash car.
(15) Mommy, you go to work?

This English-speaking child is using modal verbs to create some fairly
complex syntactic relations (e.g. 'you help me find' and 'wanna help wash'),
but some functors are missing (e.g. the article before 'car') and morphological
marking is still relatively spare, even within the confines of English (e.g. 'you
go to work?' instead of 'are you going to work?').

Finally, we offer the three longest utterances described for two children
who are at the most advanced level covered by our study. The Italian case is
a girl at 2;6 with a reported vocabulary of 614 words:

(16) Come me sta questo cappello? Papino dice che sono ridicola.
How to-me is$_{3rd sing.}$ this$_{masc.sing.}$ hat$_{masc.sing.}$? Daddy$_{3rd.sing.}$ says$_{3rd.sing.}$
that am$_{1st sing.}$ ridiculous$_{fem.sing.}$
(How is this hat on me? Daddy says that I'm silly.)

(17) Sandro faceva casino allora la mamma gli diceva 'stai zitto, se no ti do
le totte.'
Sandro$_{3rd.sing.}$ was-making$_{3rd.sing.imp.}$ uproar so the$_{fem.sing.}$
mamma$_{3rd.fem.sing.}$ to-him was-saying$_{3rd.sing.imperfect}$ 'Be$_{2nd sing. imperative}$
quiet, if not, to-you give$_{1st.sing.}$ the$_{fem.pl.}$ spankings$_{fem.pl.}$
(Sandro was making a fuss so Mommy told him 'Be quiet, if not I'll
give you a spanking.')

(18) Mamma, scriviamo una lettera a Babbo Natale, così mi porta la
bambola dei lamponi?
Mommy, let's write$_{\text{1st pl.}}$ a$_{\text{fem.sing.}}$ letter$_{\text{fem.sing.}}$ to Santa Claus$_{\text{3rd sing.}}$ so-
that to-me brings$_{\text{3rd.sing.}}$ the$_{\text{fem.sing.}}$ doll$_{\text{fem.sing.}}$ of the$_{\text{masc.pl.}}$ rasp-
berries$_{\text{masc.pl.}}$
(Mommy, let's write to Santa Claus, so he'll bring me the raspberry
doll.)

Compare these richly marked examples with the examples reported for an
American girl at 2;6 with 622 words:

(19) Morning Anna have her pants on and her pants too fat.
(20) I want to go with you to wash your car.
(21) Move soap out of here, Anna don't like it here.

These examples provide evidence for extensive use of prepositional phrases
(one of the areas where American children appeared to be slightly ahead), but
the inflectional morphology is still faulty ('Anna have' instead of 'Anna has',
and 'Anna don't' instead of 'Anna doesn't'), and at least one article is
missing.

Although these cases were selected randomly, they should not be viewed
as representative of the vast range of variation reported by parents in both
language groups. We include them here primarily because we want to
underscore that the complexity-by-vocabulary-size results in Figure 9 tell
only part of the story. In Figure 9 and in the case studies illustrated above,
it seems clear that grammatical development (as defined for that language) is
tightly yoked to lexical growth. However, these similarities mask a host of
differences in the detailed properties of grammatical morphology and syntax
that distinguish English from Italian. Among other things, Italian children
will have to acquire far more inflectional morphology than their English-
learning counterparts (Bates, 1976). This problem can be resolved in one of
two ways (with various points in between): (1) language learning may take
much longer in Italian than it does in English, or (2) Italian children may
keep pace with their English-speaking counterparts in the proportion of their
target grammar that they are able to produce at any given point. If the latter
outcome holds, then we should expect Italian children to display much more
grammatical marking than their American age mates across the period from
1;6 to 2;6. The few cases that we have reported here provide evidence for the
second option, but much more evidence will be required to settle the issue,
including evidence from free speech and structured elicitations as well as
parental report.

SUMMARY AND CONCLUSION

We set out to address two related questions regarding the passage from first words to grammar in English- and Italian-speaking children between 1;6 and 2;6:

(1) Are there cross-linguistic differences in the composition of vocabulary within and across this age range, with special reference to hypothesized differences in the onset and growth of nouns and verbs?

(2) Are there cross-linguistic differences in the pace and shape of grammatical development and its relation to vocabulary size?

Our answer to the first question is essentially 'no'. That is, we find no evidence in favour of the idea that verbs or other predicative terms get off the ground earlier in Italian, even though there are many differences in the input to small children that ought to favour verbs in that language. In fact, we actually found evidence for a small English advantage in the proportional representation of both common nouns and predicates, compared with Italian children matched for vocabulary level. However, these small differences appear to be the statistical reflex of a rather different phenomenon, namely, the fact that Italian children have a somewhat larger repertoire of social words (sound effects, names for people, social routines), particularly within the early stages when vocabularies do not exceed 200 words. Although other factors cannot be ruled out, we suspect that this particular cross-linguistic difference is the by-product of cultural differences between America and Italy, including the tendency for Italian families to live in the same city with an extended network of relatives and old family friends. In every respect, these data replicate and extend our previous findings with a separate sample of younger American and Italian children between 0;8 and 1;4 (Caselli *et al.*, 1995).

With regard to the second question, the answer depends on which aspect of the data we choose to make the point. In both languages, closed-class words are rare in the early stages of lexical development (when vocabularies are under 200 words), and subsequent growth in the function word category is tightly correlated with overall vocabulary size. However, the shape of the relationship between closed-class development and vocabulary size is somewhat different in English and Italian. Data for the American children is best fit by a nonlinear function, with little or no effect of vocabulary size under 400 words and a visible acceleration after that point. Data for the Italian children is more gradual and linear across this period of development, so that there appears to be a slight closed-class advantage for Italian in the early stages.

Switching from function word counts to a more comprehensive measure of grammatical complexity, we found a striking similarity between English and

106

Italian in the nonlinear function that ties grammatical complexity to overall vocabulary size. To some extent, this finding reflects the method that we chose to measure grammatical growth. In both languages, we constructed a 37-item scale of sentence pairs selected to reflect contrasts that are known to emerge in the period between 1;6 and 2;6, based on extensive free speech data for both languages. The items are not translations of one another, but they do reflect a similar range of morphosyntactic contrasts. When such a measure is used, the same relationship between grammatical complexity and vocabulary size appears to hold across these languages.

A qualitative comparison of the order in which specific function words are acquired also yielded more similarities than differences. Although the two lists summarized in Table 4 do not cover all the function words available in either English or Italian, they do comprise a representative list of the function words that are known to appear within this age range, for each language. Nothing was left out of these lists that might have yielded a different result from the one that we obtained. Overall, a close examination of the data in Table 4 suggests far more similarities than differences between English and Italian in the order in which specific grammatical function words are acquired (e.g. singulars before plurals in every relevant subclass; similarities in the order in which spatial locatives appear, and in person marking within the pronoun system). All of these findings are compatible with early reports by Slobin and his colleagues based on free speech and/or structured elicitations (e.g. Johnston, 1985). Some subtle and provocative differences did emerge (e.g. the subject pronoun 'I' seems to come in later in English than its Italian counterpart 'io'), suggesting that some combination of statistical and pragmatic factors can intervene to shift the acquisition point for individual items. However, given the limited nature of this data set, these small findings are best viewed as working hypotheses for research with free speech and elicitation methods that focus more narrowly on the function words in question.

Despite the host of similarities in grammatical and lexical development revealed in this study, we know *a priori* that Italian children will have to acquire a much richer array of morphological contrasts than their American counterparts. We also know that the two languages differ markedly in the kinds of word orders that children have to learn, including the many pragmatic and prosodic factors that condition the use of word order variation and subject omission in Italian. A measure that is more sensitive to these differences would undoubtedly pick up measurable differences in grammatical development in English and Italian. This prediction is supported by our brief qualitative look at the longest sentences reported for a small group of children who were randomly selected from the larger sample. When children were matched for both age and vocabulary size, sharp differences were evident in the sheer amount of grammatical morphology that Italian

children produce, reflecting the greater morphological load that they have to acquire. We think it would be very useful to continue this procedure of matching children for vocabulary size in future cross-linguistic studies, providing a clearer look at the developmental consequences of structural differences between the languages in question.

We want to end by emphasizing that these cross-linguistic results are largely compatible with findings obtained using free speech methods in the last two decades of cross-linguistic research by other investigators. We have provided some new perspectives on this issue, but the overall picture is the same. That is, language-specific variations can be observed in the content of language development, but all children must acquire their language under heavy and presumably universal constraints from perception, production, memory, and the availability of cognitive/conceptual structures that underlie all human languages.

REFERENCES

Au, T., Dapretto, M. & Song, Y.-K. (1994). Input vs. constraints: early word acquisition in Korean and English. *Journal of Memory and Language* 567–82.

Bamberg, M. G. W. (1988). *Form and function in the construction of narratives: developmental perspectives*. Berlin: Mouton de Gruyter.

Bates, E. (1976). *Language and context: the acquisition of pragmatics*. New York: Academic Press.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. (1979). *The emergence of symbols: cognition and communication in infancy*. New York: Academic Press.

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

Bates, E. & Devescovi, A. (1989). Crosslinguistic studies of sentence production. In B. MacWhinney & E. Bates (eds), *The crosslinguistic study of sentence processing* (225–53). New York: Cambridge University Press.

Bates, E. & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. In G. Altmann (ed.), Special issue on the lexicon, *Language and Cognitive Processes* **12(5/6)**, 507–86.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, S., Reilly, J. & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* **21(1)**, 85–124.

Berman, R. A. & Slobin, D. I. (1994). *Relating events in narrative: a cross-linguistic developmental study* [in collaboration with Ayhan Aksu-Koc *et al.*]. Hillsdale, NJ: Erlbaum.

Bottari, P., Cipriani, P., Pfanner, L. & Chilosi, A. M. (1993). Inferenze strutturali nell' acquisizione della morfologia libera italiana [Structural inferences in the acquisition of free Italian morphosyntax]. In E. Cresti & M. Moneglia (eds), *Ricerche sull'acquisizione dell'italiano* [Research on the acquisition of Italian]. Rome: Bulzoni.

Braine, M. D. S. (1976). Children's first word combinations (with commentary by Melissa Bowerman). *Monographs of the Society for Research in Child Development* **41**, Serial no. 164.

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

Caselli, M. C., Bates, C., Casadio, P., Fenson, L., Fenson, J., Sanderl, L. & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development* **10**, 159–99.

Caselli, M. C. & Casadio, P. (1995). *Il primo vocabolario del bambino: Guida all'uso del questionario MacArthur per la valutazione della comunicazione e del linguaggio nei primi anni di vita* [The child's first vocabulary: User's guide to the MacArthur questionnaire for the evaluation of language and communication in the first years of life]. Milan: FrancoAngeli.

Cheng, S.-W. (1994). *Beginning words of three children acquiring Mandarin Chinese*. Manuscript, National Chung Cheng University, Taiwan.

Choi, S. & Bowerman, M. (1991). Learning to express motion events in English and Korean: the influence of language-specific lexicalization patterns. *Cognition* **41(1–3)**, 83–121.

Choi, S. & Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language* **22(3)**, 497–529.

Cipriani, P., Chilosi, A. M., Bottari, P. & Pfanner, L. (1993). *L'acquisizione della morfosintassi in italiano: Fasi e processi* [The acquisition of morphosyntax in Italian: stages and processes]. Padua: Unipress.

Cresti, E. & Moneglia, M. (eds). (1993). *Ricerche sull'acquisizione dell'italiano*. [Research on the acquisition of Italian]. Rome: Bulzoni.

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Research* **34**, 565–71.

Dale, P. S. (1996, September). *Cognition and the emergence of combinatorial language: a cross-linguistic perspective*. Presented at the symposium 'The Emergence of Human Cognition and Language,' Waseda University, Tokyo, Japan.

Demuth, K. A. (1989). Maturation and the acquisition of the Sesotho passive. *Language* **65(1)**, 56–80.

Devescovi, A., Caselli, M. C. & Bonanni, R. (1996). Valutazione del primo sviluppo morfosintattico: Confronto fra due strumenti di rilevazione [Evidence of early morpho-syntactic development: a comparison between two instruments]. *Abstracts, National Congress of the Italian Psychological Association, Developmental Psychology Section*, Capri, 114–15.

Devescovi, A., D'Amico, S., Smith, S., Mimica, I. & Bates, E. (in press). The development of sentence comprehension in Italian and Serbo-Croatian: local versus distributed cues. In D. Hillert (ed.), *Sentence processing: a cross-linguistic perspective*. San Diego, CA: Academic Press.

Devescovi, A. & Pizzuto, E. (1995). Lo sviluppo grammaticale [The development of grammar]. In G. Sabbadini (ed.), *Manuale di neuropsicoligia dell' età evolutiva* [Handbook of developmental neuropsychology] (pp. 260–85). Bologna: Zanichelli.

Dromi, E. (1987). *Early lexical development*. New York: Cambridge University Press.

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

Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E., Hartung, J., Pethick, S. & Reilly, J. (1993). *The MacArthur Communicative Development Inventories: user's guide and technical manual*. San Diego, CA: Singular Publishing Group.

Gathercole, V. & Min, H. (1997). Word meaning biases or language-specific effects? Evidence from English, Spanish and Korean. *First Language* **17**, 31–56.

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

Gentner, D. (1997, August). Nouns, verbs and the division of determinism. Paper presented at the Meeting of the Northern European Language Acquisition Society, Odense, Denmark.

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

Goodman, J. C. & Nusbaum, H. C. (eds). (1994). *The development of speech perception: transitions from speech sounds to spoken words*. Cambridge, MA: MIT Press.

Gopnik, A. & Choi, S. (1990). Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language* **10**, 199–215.

Gopnik, A. & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: nouns are not always learned before verbs. In M. Tomasello & W. Merriman (eds), *Beyond names for things: young children's acquisition of verbs*. Hillsdale, NJ: Erlbaum.

Greenfield, P. & Smith, J. (1976). *The structure of communication in early language development*. New York: Academic Press.

Imai, M. & Gentner, D. (1997). A cross-linguistic study of early word meaning: universal ontology and linguistic influence. *Cognition* **62**, 169–200.

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**(**3**), 523–49.

Johnston, J. R. (1985). Cognitive prerequisites: the evidence from children learning English. In D. Slobin (ed.), *The crosslinguistic study of language acquisition* (Vol. 2). Hillsdale, NJ: Erlbaum.

Karmiloff-Smith, A. (1979). *A functional approach to child language : a study of determiners and reference*. New York: Cambridge University Press.

Kuhl, P. K. (1991). Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech. In S. Brauth, W. Hall & R. Dooling (eds), *Plasticity of development*. Cambridge, MA: MIT/Bradford Books.

Leonard, L. B., Bortolini, U., Caselli, M. C., McGregor, K. & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: the status of features in the underlying grammar. *Language Acquisition*, **2**(**2**), 151–79.

Lieven, E. & Pine, J. (1990). [Review of the book *From first words to grammar : individual differences and dissociable mechanisms*.] *Journal of Child Language* **17**, 495–501.

MacWhinney, B. & Bates, E. (eds). (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Marchman, V. & Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language* **21**, 339–66.

McDonough, L., Choi, S., Bowerman, M. & Mandler, J. (in press). The use of preferential looking as a measure of semantic development. In E. L. Bavin and D. Burnham (eds), *Advances in infancy research*. Norwood, NJ: Ablex.

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).

Nelson, K. & Lucariello, J. (1985). The development of meaning in first words. In M. D. Barrett (ed.), *Children's single-word speech*. New York: Wiley.

Ninio, A. (1993). On the fringes of the system: children's acquisition of syntactically isolated forms at the onset of speech. *First Language* **13**, 291–313.

O'Grady, W. (1987). *Principles of grammar learning*. Chicago: University of Chicago Press.

Pae, S. (1993). *Early vocabulary in Korean : are nouns easier to learn than verbs?* Unpublished doctoral dissertation, University of Kansas, Lawrence, Kansas.

Pizzuto, E. & Caselli, M. C. (1992). The acquisition of Italian morphology: implications for models of language development. *Journal of Child Language* **19**, 491–557.

Pizzuto, E. & Caselli, M. C. (1993). The acquisition of Italian morphology – a reply to Hyams. *Journal of Child Language* **20**(**3**), 707–12.

Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (ed.), *The crosslinguistic study of language acquisition* (Vol. 2, 1157–1256). Hillsdale, NJ: Erlbaum.

Slobin, D. (ed.) (1985–97). *The crosslinguistic study of language acquisition* (Vol. 1–5). Hillsdale, NJ: Erlbaum.

Tardif, T. (1996). Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. *Developmental Psychology* **32**(**3**), 492–504.

Tardif, T., Shatz, M. & Naigles, L. (1996). The influence of caregiver speech on children's use of nouns versus verbs: a comparison of English, Italian and Mandarin. Abstracts of the VIIth International Congress for the Study of Child Language, Istanbul.

Tomasello, M. (1992). *First verbs : a case study of early grammatical development*. Cambridge: C.U.P.

Volterra, V. (1979). First words in language and action: a qualitative look. In E. Bates, L. Benigni, I. Bretherton, L. Camaioni & V. Volterra (eds), *The emergence of symbols : cognition and communication in infancy* (141–82). New York: Academic Press.

Volterra, V. & Caselli, M. C. (1986). First stage of language acquisition through two

modalities in deaf and hearing children. *Italian Journal of Neurological Science, Supplement 5, Maturation and Learning*, 109–15.

Weist, R., Lyytinen, P., Wysocka, J. & Atanassova, M. (1997). The interaction of language and thought in children's language acquisition: a crosslinguistic study. *Journal of Child Language* **24**, 81–121.

Werker, J. F. (1994). Cross-language speech perception: developmental change does not involve loss. In J. C. Goodman & H. C. Nusbaum (eds), *The development of speech perception: the transition from speech sounds to spoken words* (93–120). Cambridge, MA: MIT.