J. Child Lang. **29** (2002), 735–757. © 2002 Cambridge University Press
DOI: 10.1017/S0305000902005330 Printed in the United Kingdom

Early lexical development in German: a study on vocabulary growth and vocabulary composition during the second and third year of life*

CHRISTINA KAUSCHKE
AND CHRISTOPH HOFMEISTER
Free University of Berlin

(*Received* 25 *April* 2000. *Revised* 12 *November* 2001)

ABSTRACT

This paper focuses on aspects of early lexical acquisition in German. There have been conflicting results in the literature concerning both the pattern of vocabulary growth and the composition of the early lexicon. Our study describes the development of various categories of words and questions the preponderance of nouns in spontaneous speech. 32 children were studied longitudinally through recordings made at age 1;1, 1;3, 1;9 and 3;0. The following properties of the data were investigated: vocabulary size in relation to age, frequency of word use, and distribution of word categories. The results show that use of both types and tokens increases with time. A trend analysis indicates an exponential increase in vocabulary production in the second year, followed by a further expansion. This vocabulary spurt-like pattern can be observed in the use of word types and tokens. The findings in regard to vocabulary composition illustrate the dynamics present in the development of word categories. In the beginning, children use mostly relational words, personal-social words and some onomatopoeic terms. These categories are gradually complemented with nouns, verbs, function words and other words so that we see a balanced lexicon by 3;0. Trend analyses clarify characteristic developmental patterns in regard to certain word categories. Our spontaneous speech data does not support a strong noun-bias hypothesis.

[*] We would like to thank Gisela Klann-Delius for making available the database analysed in this study and for valuable comments on the paper. We thank Inge Lasser and James Murphy for their assistance with the English version of the paper. We also thank two anonymous reviewers and the action editor for helpful comments on earlier versions of the manuscript. Address for correspondence: Christina Kauschke, Freie Universität Berlin, Dept. of Philosophy and Humanities, German and Dutch Languages and Literature, Habelschwerdter Allee 45, 14195 Berlin, Germany. Tel: +49-30-838 54808; e-mail: kauschke@zedat.fu-berlin.de

INTRODUCTION

How children acquire and use words is one of the central issues of language acquisition research. The acquisition of the lexicon is seen as one separate and essential part of the acquisition process. Research and discussions have proceeded from different perspectives. For example, the course of lexical acquisition has been described on the basis of a variety of empirical findings, which in some cases are incompatible with each other. Descriptions of lexical acquisition have been given in terms of lexicon size and growth patterns as a function of age. Another controversial subject has been the composition of the infant lexicon.

Most of the findings have been limited to the English language, so that any conclusions on early lexical acquisition in other languages are premature due to the lack of systematic empirical studies on large enough corpora. As for German, there are only a few studies on lexical acquisition and they focus on different issues. Stern & Stern (1928) and Elsen (1999) remark on early vocabulary development in individual children. Cross-sectional studies have typically been concerned with late preschoolers (Augst, 1984) or pupils in elementary school (Pregel & Rickheit, 1987).

In this paper we will first provide a survey of empirical findings on the course of early vocabulary development and the composition of the lexicon. Then we will present the results of a study on the acquisition of the lexicon in 32 two- to three-year-old German children. These results focus on vocabulary growth and the composition of the lexicon. We aim to show which growth patterns occur in spontaneous word production and we will trace the development of different word categories in German. We will also address the issue of whether nouns are the dominant word category in spontaneous production, as has sometimes been suggested.

Vocabulary growth

The occurrence of the first words constitutes an important milestone in infant language acquisition. Bloom, Tinker & Margulis (1993) have determined the average age for the production of the first words at 1;1·26 with a range from 0;10·5 to 1;5·23. Although individual children have been reported to produce their first words at the age of 0;8, it is generally around the first birthday that they begin their productive use of referential words (in the case of undisturbed development). After that point, the size of their vocabulary initially grows rather slowly. Around the age of 1;6–1;7 the productive vocabulary of most children contains roughly 50 words. For instance, Menyuk, Liebergott & Schultz (1995), studying 53 children, found an average age of 1;6·4 for the '50-word-benchmark', with a range between 1;3·8 and 1;8·9. In many cases this is the time period in which a sudden

growth of the vocabulary can be observed. This has been shown in numerous studies and is referred to as THE VOCABULARY SPURT. It is clear that determination of vocabulary size (e.g. 50 words) at a certain age establishes a static quantitative figure, while the concept of vocabulary spurt takes into account the dimension of time and is thus indicative of the dynamics of vocabulary acquisition.

Using the criterion 'at least three new words per week or twelve new words a month', Bloom *et al*. (1993) determined that the 14 children they examined went through the vocabulary spurt at an average age of 1;7·7, between-subject variation ranging from 1;3·2 to 2;1·6. Vocabulary growth can be quite startling in some cases, as in the example of a boy who shows relatively advanced linguistic development for his age. At an age of 1;7 he acquired 83 new words in one week (Robinson & Mervis, 1998).

Whether the vocabulary spurt is a hallmark of the development of every child is unclear, as the controversy between Goldfield & Reznick (1990, 1996) and Mervis & Bertrand (1995) reveals. Goldfield & Reznick (1990) describe children whose vocabulary increases suddenly as well as children whose lexicon develops gradually up to a vocabulary of 99 words (non-spurters). In contrast, Mervis & Bertrand (1995) see the spurt as a pattern which is present in all normally developing children. As they were able to show that a dramatic increase in vocabulary size may take place later than had been previously assumed (average age 1;8 with an average lexicon size of 112 words), they claim that the spurt occurs universally in all children, but with variable onset (regular vs. late spurters). Anisfeld, Gasparini, Hoberman & Rosenberg (1998) also found later phases of increased vocabulary growth in four children.

In looking at single cases as well as larger samples, varying patterns of acceleration with regard to vocabulary acquisition in the second and third year of life have been found:

(a) Rapid and sudden growth (vocabulary spurt/explosion), e.g. Goldfield & Reznick (1990), Bloom (1993), Robinson & Mervis (1998), Dromi (1999).
(b) Alternating pattern of more or less extended spurt intervals and plateaus, (staircase-shaped curve), e.g. Clark (1993), Menyuk *et al*. (1995), Goldfield & Reznick (1996), Anisfeld *et al*. (1998), Robinson & Mervis (1998).
(c) Extended spurt phase (long-term vocabulary surge), Goldfield & Reznick (1990).
(d) Exponential growth, e.g. Bates, Dale & Thal (1995).
(e) Gradual and linear growth, e.g. Goldfield & Reznick (1990), Bloom (1993), Fenson, Dale, Reznick, Bates, Thal & Pethick (1994).

It cannot be excluded as a possibility that the differences in the findings are due to different methods of sampling and interpretation. It is particularly important to take into consideration the complete time period during which

samples were collected as well as the temporal distance between samples. The longer the intervals between the points of measurement, the more apparent are the changes in speed of acquisition. In order to identify predominant patterns of vocabulary growth as well as the overall spectrum of variation, additional longitudinal studies with a closer sequence of sampling points are still needed.

Composition of the lexicon

Much empirical research has set out to determine which word categories are present and/or dominant in the lexicon of infants. Proponents of the so-called NOUN-BIAS HYPOTHESIS assume that nouns are the dominant word category during early lexical acquisition. However, there are diverging views on what the noun-bias hypothesis actually says:

(a) Nouns are acquired earlier than verbs / other word classes.
(b) Nouns form the majority of children's early vocabulary.
(c) Nouns in children's early vocabulary are predominantly object labels.
(d) A preference for nouns promotes further language development.

We will now examine these claims one by one. First, we will look at empirical findings concerning the developmental sequence of word categories. Then, the proportion and reference of nouns will be discussed, and finally we will address the influence of acquisition styles on vocabulary composition.

Concerning the developmental sequence of word categories (claim a), Gopnik (1988), studying the order of acquisition of social words, nouns, and cognitive-relational words in English, showed that social words with interactive function (such as 'no' or 'hello'), as well as nouns, already appear during the earliest phases of language acquisition, while relational terms come in at a later point. Gopnik stresses that most of her subjects had produced non-nominal expressions before they named any object. Gentner (1981, 1982) reports a developmental sequence in which nouns were first, followed by relational and expressive words and later by verbs. This temporal asymmetry with respect to order of acquisition is said to be valid across different languages. Gentner assumes that children generally acquire verbs only after they have acquired a certain repertoire of nouns. This assumption was partly confirmed by Caselli, Bates, Casadio, J. Fenson & L. Fenson, Sanderl & Weir (1995) for English and Italian: they argue for a universal sequence from routines to nouns to verbs. According to Gentner's account, nouns are acquired early and easily because they are formally and conceptually less complex than verbs. Nouns, especially words for objects, are said to be more concrete and more basic and are therefore grasped more easily by children. While in this view, nouns correspond to cohesive perceptual entities, verbs

are perceptually less constrained and more ambiguous. The correspondence between words and perceptual categories is thought to be considerably looser in the case of verbs.

However, recent crosslinguistic studies challenge the notion of a universal noun–verb asynchrony. Bassano (2000) noted that verbs appeared as early as nouns in a French-speaking child, although nouns predominated over verbs until age 2;0, when verb frequency gradually exceeded noun frequency. Results by Tardif (1996), Brown (1998) and Choi (1998) support the observation that the order of appearance of nouns and verbs depends on characteristics of the language acquired. Children learning Mandarin, Korean and Tzeltal, a Mayan language, produce a substantial number of verbs in early phases of their lexical development. In these studies, verbs have been found to be roughly equivalent to nouns or to even outnumber them. The verbs are morphologically marked and appear to be part of the productive vocabulary occurring in early multi-word utterances. The verb lexicon grows at an earlier point than in English, and the proportion of verbs is higher. Neither do nouns appear earlier, nor do they dominate in the early lexicon. It has been suggested that the early use of verbs in these languages is linked to features of the input, such as frequency (e.g. the number of types and tokens of nouns and verbs in child-directed speech) and saliency of verbs (e.g. the position of nouns and verbs in utterances), or to the pragmatic focus of the caregiver's utterances (e.g. talking about objects or actions) (Tardif, Shatz & Naigles, 1997; Choi, 2000; Kim, McGregor & Thompson, 2000). According to these parameters child-directed speech of Mandarin and Korean adults apparently highlights verbs more than nouns by producing verbs more frequently than nouns, by placing verbs in utterance-final position, and by focusing on action-oriented activities.

Smith & Sachs (1990) investigated the comprehension and production of verbs in English-speaking children. They characterized verb production as moderately but steadily increasing with age. Their findings suggest that verbs do not play a substantial role in early vocabulary. However, at a later point children build up and extend their verb vocabulary. Behrens (1998) studied the production of verbs in longitudinal corpora of ten children who acquired German, English and Dutch. She showed that various verbs appeared in the productive lexicon of all of the children acquiring Germanic languages especially during the third and fourth year of life. Among these verbs were also particle verbs, where the particle can be separated from the verbal stem (e.g. *zudecken*, 'to cover up') and verbs with inseparable prefix (e.g. *bedecken*, 'to cover').

With respect to the proportion of nouns in early vocabulary (claim b), some studies on lexical acquisition in English suggest that nouns amount to one third. Nelson, Hampson & Kessler Shaw (1993) found that nouns make up 38% of all words used at age 1;8. In a longitudinal study Bloom *et al.*

(1993) found that nouns which refer to objects make up about 30% of word types and tokens during the stage of 'first words' and the '50 words' stage. Lieven, Pine and colleagues have used a methodology combining interview and observation techniques for evaluating the composition of the lexicon when it contains 50 words and 100 words. An early study by Lieven, Pine & Dresner Barnes (1992) determined the proportion of nouns at 33·2% at 50 words and 37·5% at 100 words. Pine, Lieven & Rowland (1997), examining 26 children, found the proportion of nouns to be slightly higher at 42·5% and 46·1%.

Bates, Marchman, Thal, Fenson, Dale, Reznick, Reilly & Hartung (1994) conducted an extensive cross-sectional study in order to determine the time of appearance and the proportion of different word classes. The method they used was vocabulary checklists. The children's ages were 1;4 to 2;6. While the overall vocabulary during the time period examined grew constantly, the different word categories made up varying proportions at different points in time. Three 'waves' of internal reorganization were determined. After a notable increase of the proportion of nouns at an early stage, the proportion of predicates (verbs and adjectives) grew steadily. Finally the proportion of function words expanded markedly. Pine *et al*. (1997) found similar trends. During the time in which the vocabulary grew from 50 to 100 words, there was a significant increase in nouns, the class of 'other' words (which included verbs) and frozen phrases. During the same period, onomatopoeic terms, proper names and interactive terms (words with social or pragmatic function) decreased significantly.

Turning to the reference of nouns (claim c), Snyder, Bates & Bretherton (1981) and Gentner (1982) stated that nouns in children's early lexicons mainly refer to small and concrete objects. The tight correspondence between objects and nouns was thought to facilitate concept formation and word learning. Nelson *et al*. (1993), however, found that nouns in early vocabularies might refer to a variety of concepts, not all of which can be classified as representations of objects. About 40% of the nouns the children produced did not refer to basic-level objects, but to generic terms, events, locations and so forth. Thus, any noun preference in English-speaking children is only partly due to the acquisition of object names.

The issue of whether stylistic differences in individual children are related to different abilities (claim d) has been pursued ever since Nelson (1973) introduced the referential–expressive dichotomy. In a cross-sectional study, Snyder *et al*. (1981) found positive correlations between vocabulary size and a high proportion of nouns at 1;1. Bates, Bretherton & Snyder (1988) originally considered an early noun preference as an indicator of advanced linguistic skills. However, recent studies challenge the view that a noun preference facilitates further language development. In Bates *et al*. (1994,

1995) and Pine *et al*. (1997) no relation could be found between referential style ('nouniness') at the beginning of lexical development and later linguistic abilities. Lieven *et al*. (1992) do not consider the referential style a superior strategy. Instead, there are alternative ways to acquire language, which are just as effective.

In summary, alongside the hypothesis that children initially prefer nouns and predominantly refer to objects, we find proponents of the view that the composition of early vocabularies is more heterogeneous. The strongest version of the noun-bias hypothesis incorporates the characteristics (a–d) mentioned above. The literature reviewed here casts doubt on a strong hypothesis in demonstrating that nouns are not a universally preferred category during early lexical acquisition. Rather, noun preference – where it is observed – can be related to characteristics of the specific language and to influences of the pragmatic focus of child-directed speech in certain languages. In English, nouns are often acquired earlier than verbs, but they are not the first word category to appear. Relational, expressive or interactive words are equally important categories of words in the early lexicon of infants. Developmental patterns of Asian languages show that verbs are not necessarily preceded by nouns, but are rather acquired simultaneously. Not even for English do nouns constitute an overwhelming quantitative majority. A relatively large increase at the beginning of lexical acquisition of English is characteristic. Any preference for nouns beyond the early stage does not facilitate further language development.

From this we conclude that the role played by nouns during early vocabulary development does not seem to be as prominent as has been assumed. Gentner claims that a close noun–object link serves as an entry point into language. However, given that children's early nouns do not exclusively refer to objects and that children encode actions and change of states from the outset using relational words (such as 'up', 'down' etc.), her hypothesis of a preexisting perceptual and conceptual distinction, reflected in the distinction between nouns and verbs, may not be sufficient to explain early lexical acquisition. Some authors (e.g. Gopnik & Choi, 1995; Whong-Barr, 2000) have gone so far as to consider any noun preference as an artefact of the methodologies commonly used for examining the vocabulary of young children. They argue that language-specific as well as methodological factors have not been sufficiently considered so far. For instance, if vocabulary checklists are used, the proportion of nouns is expected to be higher than for an analysis of spontaneous production data (see Bornstein & Haynes, 1998 for a discussion of different methods). Despite these objections it seems reasonable to test a weaker version of the noun-bias hypothesis which claims that nouns may have priority over verbs in the developmental sequence in certain languages.

In the study discussed below we will address specific issues concerning the course of vocabulary growth and the composition of the productive lexicon in German.[1] More specifically we will deal with

(a) the size of the vocabulary and the frequency of word use in relation to age (development of types and tokens)
(b) growth rates: Is a pattern of vocabulary spurt to be found in spontaneous word production?
(c) the composition of the lexicon, e.g. the proportion of word categories: Which categories are used at specific ages? Is there any evidence supporting the claim that nouns play a privileged role at all?

For such purposes, spontaneous production data, as used in the study to be reported, forms a suitable database because it allows the researcher to determine the overall frequency of words as well as to establish which proportion of the total vocabulary is made up of a particular word category. This cannot be done with checklists.

METHOD

Subjects

32 children (16 boys, 16 girls) were studied during their second and third year of life.[2] Information about the subjects' socio-economic status was gained through a questionnaire assessing the education and current profession of the parents. Three quarters of the children came from middle-class or upper middle-class families; one quarter came from lower middle-class or working-class families. Most children were first-borns. All children had gone through normal birth and were developing normally during their first year of life. The longitudinal study comprised four recording samples, three of which were taken during the second year, at age 1;1, 1;3, and 1;9. The final recording was taken at the age of 3;0. This relatively late sampling point was intended as a point of comparison.

During each 30-minute recording session the mother interacted freely with her child in a room furnished with appropriate toys. Ten minutes of these videotaped sessions were included in the analysis where all verbal utterances,

[1] The study presented here forms part of a larger investigation which has dealt with various aspects of lexical acquisition in German (Kauschke, 1999, 2000). This larger study documented spontaneous word production during the second and third year of life. In addition, the predictive power of individual differences with respect to vocabulary size for later development of language was examined.

[2] The data was taken from the extensive database collected for the project 'Emotional quality of mother–child interaction and its impact on the development of communicative competence in the child', directed by G. Klann-Delius at the Free University of Berlin (see Klann-Delius, Hédervári & Hofmeister, 1996). This project was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within its focus program 'Language Acquisition', and by the Köhler foundation.

as well as vocal expressions (vocalizations, babbling), other sounds (e.g. crying, sniffing), and gestures of the mother and child were transcribed in chronological order (see Klann-Delius, 1990 for details on this method). These transcripts formed the basis for all the vocabulary analyses.

Procedure

Determination of word status. The transcripts contained utterances of various expressive modes (see above). To extract the words, a coding procedure was employed that had already been established and used in a larger research project (see Klann-Delius, Hédervári & Hofmeister, 1996).[3] Among many other parameters, the symbolic status of all utterances had been determined by regarding the whole situational and communicative context provided in the videotapes. The coding scheme distinguished

(a) gestures,
(b) vocalizations (babbling),
(c) paralinguistic expressions (sounds like coughing),
(d) expressive utterances (vocal expressions like laughing or crying; expressive interjections),
(e) speech (conventional words, 'speech-like' word forms with phonological simplifications which are typical for the early words of children, short utterances consisting of discourse particles like 'yes' and onomatopoeic forms).

We looked for word candidates in all utterances, regardless of length, including onomatopoeic forms and conventional expressive interjections (e.g. 'oh'). In order to differentiate words from babbling it is necessary to take the situational context into consideration. An example may illustrate this problem: the utterance |lɪli| may appear as reduplicated babbling during play. In another situation, a child may use this form as an object name, pronounced with reduced phonological complexity (age-appropriate simplification of the target item *Brille* – 'eye-glasses'). The utterances finally chosen as words met the conditions proposed by Vihman & McCune (1994): phonetic consistency, semantic content and use in a plausible context.

Every word determined in this fashion was then noted down in citation form (not in its actual realization); morphological markers were not taken into account. Every word type and every word token was stored in a database which had been developed especially for this purpose. Within this database, the words were sorted into lexical categories (see below). The final database consisted of 751 different words which were used by the children 3440 times when calculated in terms of types, and 9115 times when calculated in terms of tokens.

[3] For details about the coding procedure see Klann-Delius (1992).

Coding of word categories. The next step consisted in the assignment of words to lexical categories. As Bloom *et al*. (1993: 432) have pointed out, assigning words to word classes in early vocabularies constitutes a serious methodological problem for empirical studies on lexical acquisition. Especially during the one-word stage, syntactic and distributional evidence is not always available, so that ambiguities are inevitable. Furthermore, early word meanings are unstable in the beginning, so that category membership of a word may change over time. Numerous classification systems have been considered in the literature. They differ in the number of categories assumed as well as in the category types and definitions. These classifications also differ in the manner in which words are assigned to categories. According to Dromi (1987), classification systems serve different functions: they may be used to determine psychologically real categories, to reflect communicative intentions, or they may be applied as methodological means for data reduction.

The highly differentiated coding system in Dromi (1987), although very useful, is rather time-consuming and therefore only applicable to small samples. The classification system used here is intended as a means of structuring and organizing the data with the aims of providing comparable codings for each of the 32 subjects and of detecting developmental trends. Our classification system consists of nine main categories to which words were assigned (see Table 1). The categories were established during a pilot phase with a subset of the transcripts. After refining the categories, the coding scheme was applied to the overall database. In assigning words to one of the nine categories, two criteria were taken into account: a word had to resemble an adult word form, and the category to which the word was assigned had to be compatible with the linguistic context (the child's utterance in which the word was embedded; previous and following utterances of mother or child).[4] In short, the procedure chosen here allows for the comparison of word categories between subjects, but might conceal idiosyncrasies in word use.

Statistical analysis

The frequencies of types and tokens were counted per child and per recording. This data was then entered into the software package SPSS. Basic statistics were performed by contingency tables and – where appropriate – analysis of variance.

To measure changes in vocabulary size (number of types and tokens) over time, a one-factorial analysis of variance (ANOVA) with repeated measures

[4] Unlike English, German has almost no homophones resulting from conversion that can be either nouns or verbs (compare, for example, the words 'the hammer' and 'to hammer' in English to 'der Hammer' and 'hämmern' in German). For this reason, the number of ambiguous words turned out to be relatively low.

TABLE 1. *Classification of word categories*

Category	Category members	German examples	Translation of examples
nouns	proper names, persons, animate and inanimate objects, abstract nouns, internal state-nouns	Holger	
		Baby	baby
		Ball	ball
		Idee	idea
		Angst	fear
verbs – action words	actions and activities with and without objects, events, movements, states, mental verbs, internal-state verbs	suchen	to search
		rennen	to run
		weinen,	to cry
		glauben	to believe
adjectives and modifiers	adjectives denoting attributes or qualities, modifying elements, internal-state adjectives	heiß, vier	hot, four
		böse, toll	bad, fantastic
personal-social words	interactive words like assertions, discourse signals and attention-getting devices; personal-expressive utterances, interjections	ja, nein	yes, no
		hallo, danke	hello, thanks
		hm	
		guck	look
		aua	ouch
relational words	words for relations between actions and/or objects, words for appearance/disappearance, words for functions of objects, words for temporal aspects of actions or events, locatives	da, weg	there, all gone
		oben	up
		ran, auf	on
		wieder	again
pronouns	personal pronouns, demonstrative pronouns, possessive pronouns etc.	du	you
		dies	this
		sein	his
function words	prepositions, auxiliaries, articles, conjunctions, question words etc.	aus	off
		haben	have
		ein	a
		weil	because
		warum	why
onomatopoeic terms	onomatopoeic utterances	brumm,	car sound
		tatütata	fire engine sound
other words	particles; words with ambiguous status	eben, denn	just

was applied to the data ($n = 32$). A polynomial decomposition of the data measured at different points in time was performed. This yielded estimates of linear, quadratic and cubic trends. Since the logarithm of an exponential growth is a straight line, the analysis was based on the logarithms of the data instead of the data itself. Therefore, a significant linear trend was interpreted as an indicator of exponential change over time.

Prior to analysis of word categories, the data was converted to relative frequency by dividing the number of types and tokens of every category by the respective totals (sum of all types or tokens for every child). If a child did

not utter at least one type (and token) at a certain point in time this transformation was not applied. Unfortunately, this led to the exclusion of some children, reducing the sample to $n = 28$ (for the calculations on word categories only). As a result of the conversion to proportions, the data did not fit the assumptions of parametric methods because these methods request an analysis of the variable on an interval scale and they assume a normal distribution of population parameters. Employing parametric methods in this context would thus lead to floor and ceiling effects and possibly to distortions of the results. Instead of a parametric test, the Friedman test was used (Friedman, 1937). This test makes fewer assumptions, inasmuch as the measurement of the variable is only required to be on an ordinal scale and no assumptions about the distribution of population parameters are made. This means that the computation of the Friedman test just takes into account the order of the values, ignoring distances in between values. There is some loss of information compared to measurements on an interval scale, but if the calculation is based on exact probabilities, the Friedman test proves to be even more powerful than a parametric test. Changes in the proportion of word categories over time were therefore measured by trend tests derived from the Friedman test. This procedure involved performing a polynomial decomposition of the data measured at different points in time. The decomposition yielded estimates of linear, quadratic and cubic trends. Since the relative frequencies submitted to analysis were not derived independently, the Bonferroni correction procedure was performed on the *p*-values yielded by the trend tests.

RESULTS

Vocabulary growth

As expected, the number of words produced by the 32 children during each 10-minute recording session increases with age. Table 2 shows an overview

TABLE 2. *Descriptive values* (%) *for types and tokens*, n = 32

Age	13 months	15 months	21 months	36 months
Total number of types	82	187	649	2522
Mean types	2·56	5·84	20·28	78·81
Range types	0–10	0–16	0–54	10–138
S.D. types	2·18	3·8	12·17	31·79
Total number of tokens	302	556	1846	6411
Mean tokens	9·44	17·38	57·69	200·34
Range tokens	0–52	0–58	0–140	18–485
S.D. tokens	10·66	12·89	36·00	104·18
Type-token-ratio	0·44	0·40	0·38	0·42

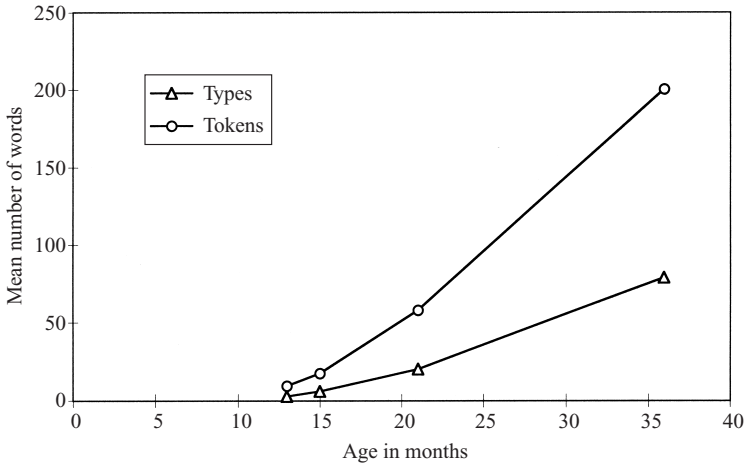

Fig. 1. Growth patterns of types and tokens, $n = 32$.

of the total and mean numbers and the distribution in the production of types and tokens.

Growth of word types. The growth rate for the types ranges from an average of 2·5 to 78·8 and constitutes a significant increase as a factor of age ($F(3{\cdot}93) = 169{\cdot}36$, $p < 0{\cdot}001$). The variable 'age' is responsible for the variance in types to a high degree (84%). Thus, despite individual differences, general developmental sequences are evident.

Figure 1 illustrates the growth pattern of the types and tokens. From age 1;1 to 1;9 the use of different words increases in a nonlinear fashion. A trend analysis of the logarithms of the type values during the second year suggests exponential vocabulary growth (linear trend: $F(1{\cdot}27) = 239{\cdot}77$, $p < 0{\cdot}001$). This trend does not continue into the third year, however. In the analysis of all four sampling points, also the quadratic trend is significant ($F(1{\cdot}27) = 66{\cdot}99$, $p < 0{\cdot}001$). In Figure 2, the course of the trend graphs is shown using logarithms of the values. These graphs show that vocabulary growth levels out at the final sampling point (3;0).

Growth of word tokens. The frequency with which the children use a given word likewise increases with age. Table 1 shows how mean values grow from 9 to 200 words. Again the factor 'age' is statistically relevant ($F(3{\cdot}93) = 95{\cdot}24$, $p < = 0{\cdot}001$). Age explains 75% of the token variance. Thus the frequency of word use is predicted by age. The growth pattern of the tokens is comparable with that of the types (see Figure 1 and 2). The results of the trend analyses for token values confirm an exponential component during the

https://doi.org/10.1017/S0305000902005330 Published online by Cambridge University Press

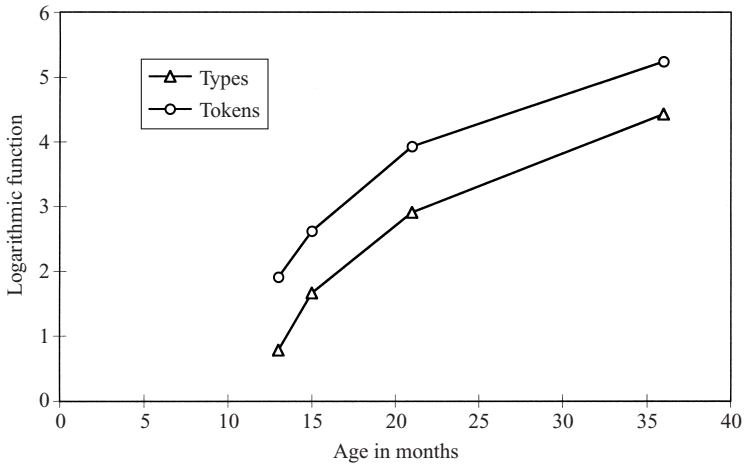

Fig. 2. Growth patterns of types and tokens (trends), $n = 32$.

second year (linear trend of logarithm values from age 1;1 to 1;9: $F(1,27) = 114{\cdot}32$, $p < 0{\cdot}001$) with a subsequent slow-down of the growth rate (quadratic trend at four measuring points: $F(1{\cdot}27) = 26{\cdot}77$, $p < 0{\cdot}001$).

Vocabulary composition

Table 3 shows the mean proportions of each word category relative to all types and tokens at age 1;1, 1;3, 1;9 and 3;0.

As illustrated, the composition of the children's vocabulary changes with age. At the outset relational words and personal-social words are predominant. At the first sampling point, they constitute more than three quarters of the lexicon (types). Many of these words are used by all children. An example of a relational word that shows consistent use across children is the word *da* ('there'). Obviously, the single word *da* with its multiple functions constitutes a dominant and important element in German children's early language acquisition.[5] *Da* is used at the first sampling point by 26 children. In contrast, there is no noun which is used by more than four children at this time. The overall proportion of relational and social words in the composition of the

[5] The word *da* ('there') which makes up a very high portion of all types and tokens at age 1;1 and 1;3 was consistently classified as a relational word in this study. Certainly, different functions for this word can be determined. In the first place, *da* appears as a relational word in a narrow sense, when the child refers to the appearance of objects. Secondly, many children use the one-word utterance *da* with deictic function to draw the mother's attention to an interesting object. Thus, as an alternative, it would have been possible in some cases to classify *da* as a personal-social word. In any case, the child would have first had to recognize the presence of the object. Once multi-word combinations appear, *da* can also be analysed as a pronominal placeholder helping to enlarge and differentiate syntactic structures (Tracy, 1991).

TABLE 3. *Proportion of word categories according to analysis of variance* (n = *28*)*; means and standard deviations for types and tokens*

	Mean	S.D.	Mean	S.D.	Mean	S.D.
	Nouns		Verbs		Adjectives	
Types						
13 months	5·83	11·70	0·00	0·00	2·38	9·85
15 months	13·81	14·83	2·87	6·73	3·42	8·41
21 months	25·10	14·95	9·04	8·69	3·58	4·90
36 months	16·09	5·20	23·57	4·06	6·04	3·09
Tokens						
13 months	4·05	10·32	0·00	0·00	2·18	10·64
15 months	10·39	14·35	1·28	3·31	2·22	5·46
21 months	21·51	14·48	7·99	8·83	3·13	4·91
36 months	11·41	5·54	21·76	4·94	3·24	1·88
	Personal-social words		Relational words		Onomatopoeic words	
Types						
13 months	34·02	29·56	50·63	35·09	4·91	10·94
15 months	35·94	20·58	26·84	21·24	7·67	11·69
21 months	22·66	12·81	23·90	8·70	5·74	9·00
36 months	7·58	2·62	16·66	4·50	0·60	0·93
Tokens						
13 months	25·19	28·07	62·45	34·22	2·88	7·55
15 months	27·14	23·39	43·57	31·56	10·88	19·29
21 months	18·99	9·32	36·50	16·14	4·94	9·02
36 months	9·36	2·85	18·75	5·49	0·40	0·71
	Pronouns		Function words		Other words	
Types						
13 months	2·23	6·85	0·00	0·00	0·00	0·00
15 months	5·78	9·67	1·38	4·25	2·30	6·05
21 months	5·10	6·33	2·68	3·29	2·23	2·96
36 months	11·58	3·58	11·88	3·45	6·02	2·10
Tokens						
13 months	3·25	12·84	0·00	0·00	0·00	0·00
15 months	2·56	4·46	0·93	3·15	1·04	3·10
21 months	4·52	7·25	1·25	1·73	1·16	1·77
36 months	15·11	4·33	12·83	2·81	7·14	3·06

type set and token set, respectively, decreases continuously. The proportion of the relational words first shows a strong decrease, followed by a less drastic reduction.

From the outset, nouns are present and their proportion grows in particular during the second year of life. While at the first and second sampling point (i.e. 1;1 and 1;3) nouns refer mainly to persons and basic-level objects, as of age 1;9 they also refer to subordinate objects and later on

TABLE 4. *Trend analysis for each word category* (n = 28)

Trend	Tokens *p*	Types *p*	Tokens *p*	Types *p*	Tokens *p*	Types *p*
	Adjectives		Verbs		Nouns	
linear	0·000001**	0·000002**	< 0·000001**	< 0·000001**	0·000341*	0·010167
quadratic	0·144461	0·224117	0·204938	0·165185	0·000003**	0·000240*
cubic	0·462877	0·474932	0·309476	0·435988	0·494227	0·420714
	Personal-social words		Relational words		Pronouns	
linear	0·000557*	< 0·000001**	< 0·000001**	0·000004**	< 0·000001**	< 0·000001**
quadratic	0·021389	0·230247	0·208702	0·111003	0·467605	0·488712
cubic	0·256718	0·025118	0·127026	0·007152	0·417886	0·188460
	Other words		Function words		Onomatopoeic	
linear	< 0·000001**	< 0·000001**	< 0·000001**	< 0·000001**	0·349495	0·308391
quadratic	0·348908	0·408264	0·388557	0·407752	0·020069	0·046261
cubic	0·372762	0·314354	0·474816	0·480502	0·209574	0·337069

*, $p < 0·05$ after Bonferroni's correction.
**, $p < 0·01$ after Bonferroni's correction.

to abstract entities. The use of compound and derived nouns starts at age 1;9. At 3;0 the proportion of nouns is below 25% in all children. At the age of 1;3 verbs begin to appear. Their proportion increases considerably until they constitute the largest proportion at age 3;0. Adjectives are also present from the outset. Their proportion remains relatively small throughout the period studied. Function words begin to emerge at 1;3 and they increase over time.

Course of development with respect to lexical categories

The issue of whether the differences in the mean proportions of the lexical categories are significant at the four sampling points was tested using the Friedman ranking test. This test indicated that the average proportion of individual lexical categories does not remain constant during the course of development. Onomatopoeic terms form the only exception to this, as their proportion does not undergo a significant change. In a second step, trend tests were carried out in order to investigate whether the changes in proportion follow any particular direction or pattern. The results in Table 4 show that all lexical categories (with the exception of onomatopoeic terms) exhibit linear trends. This is true for the composition of the type set as well as the token set. Verbs, adjectives, pronouns, function words and the category of 'other' words (comprising mainly particles) exhibit linear increasing trends. The proportions of personal-social words and relational words show linear decreasing trends.

A particular pattern is exhibited by nouns. This category demonstrates a clear non-linear trend, showing that the proportion of nouns grows in a non-linear fashion. The means of the proportions of nouns follow a significant quadratic trend ($p = 0{\cdot}00024$ for types; $p = 0{\cdot}000003$ for tokens). After an initial growth phase, the proportion of nouns reaches a peak at 1;9. At that time nouns constitute 25% of the types and 21·5% of the tokens. After that point they drop again. The graphs in Figure 3 show developmental patterns for several lexical categories.

A comparison of the composition of the lexicon for types and tokens indicates considerable overlap for most lexical categories (compare Table 3). Taking frequency into account, we see that relational words stand out by their high frequency during the second year. However, for nouns the proportion of types is above the proportion of tokens at every point in time.

DISCUSSION

Empirically supported claims regarding the development of children's vocabulary have so far been largely restricted to the English language. Moreover, for the most part they have been based on vocabulary checklists. The results presented here on the spontaneous use of words in 32 German children both confirm and supplement the existing picture. The period discussed here covers the occurrence of the first words (at age 1;1), the

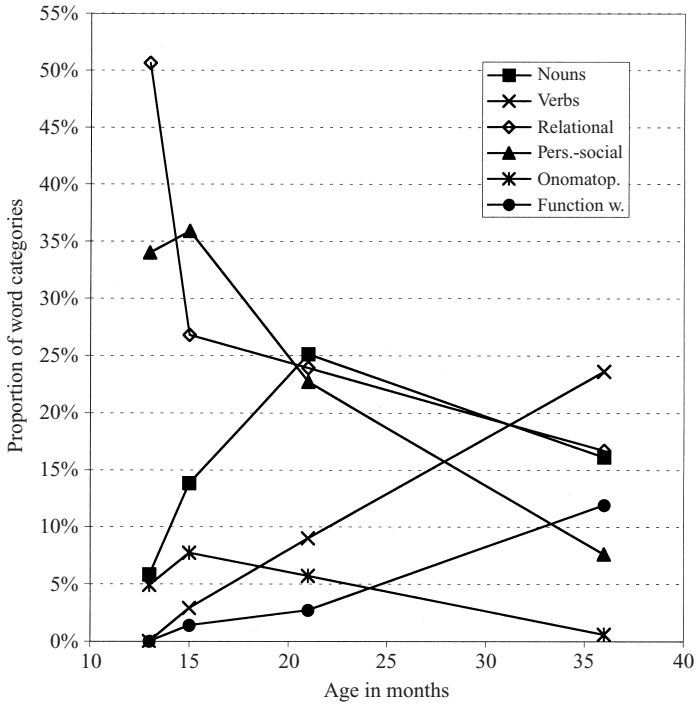

Fig. 3. Development of word categories (trends), $n = 28$.

manifest growth of the vocabulary (age 1;3 to 1;9) and the development thereafter (up to age 3;0). Looking at the spontaneous use of words, we see that the diversity of words increases along with the overall number, first exponentially and then less so, but still considerably. It is in this sense that the acceleration in the developmental curve supports a vocabulary spurt in the second year of life. This pattern has been shown for absolute vocabulary size (measured by checklists) in previous studies and it has now also been demonstrated for a number of words actually produced within a limited time period. The developmental patterns for types and tokens show an exponentially increasing curve, i.e. there is neither a particular turning point (as expected for a classic 'explosion'), nor is there continuous linear growth. Further generalizations will have to await studies incorporating additional sampling points during the third year of life. Nevertheless, current results confirm the findings of Bates *et al.*, who have described lexical development within a 'region of acceleration' (Bates *et al.*, 1995: 105).

Although the findings clearly support a general trend of a vocabulary spurt in the second year, individual patterns may differ from the group results. When we examine the number of types and tokens identified for individual

subjects, we may well find discontinuous phases or plateaus in lexical development, which supports the notion of various possible profiles of lexical growth in individual children.[6]

Children's lexical progress is not only characterized by increasing numbers of words, but also by an expansion in the number of different lexical categories which they have at their disposal. The group results for German children clearly show that it is not the nouns but the relational words as well as the personal-social words which dominate in children's productions at the early stages of development. During their second and third year of life children keep adding new lexical categories to their lexicon. As a result, word categories which are initially dominant undergo a relative decline. In the first instance, it is the nouns which expand. At no time in the samples did the nouns constitute the majority. The highest proportion of nouns was observed at age 1;9. At age 3;0 the lexicon is balanced, with no lexical category taking up more than 25%. In comparison with the above-mentioned studies on English-speaking children, the proportion of nouns is slightly smaller. Gopnik & Choi (1995) and Pine, Lieven & Rowland (1996) have pointed out that nouns tend to be overestimated in parents' reports. Thus, as the results of the present study are based on observational measures, the proportion of all nouns available might in fact be underestimated here, because the children may produce only a limited number of the nouns they have already acquired. As Pine *et al*. (1996: 586) state, the 'true value' may be somewhere between the checklist and the observational scores.

Concerning the differences in the proportions of the word categories in the type set and the token set, a high frequency of use is shown to exist during the second year for the relational words, both in terms of types and in terms of tokens. The noun proportions for types and tokens confirm the findings of Gopnik & Choi (1995), who observed that the proportion of nouns in terms of types is expected to be higher than in terms of tokens. When we look at the composition of the entire German lexicon at three years, we find that the composition of the type set does not differ considerably from the composition of the token set.

Trend analyses of the course of development of individual word categories reveal complementary developmental processes. First, there is a significant decrease in those categories which are predominant initially. Conversely, those categories which are added at a later point increase accordingly. In accordance with the findings of Bates *et al*. (1994) and Pine *et al*. (1997), some lexical categories follow characteristic developmental curves during particular periods in development. After children have acquired relational, social as well as some onomatopoeic words as their first lexical items, these

[6] For details on the language development of three single cases with contrasting linguistic abilities see Kauschke (2000).

are increasingly replaced by other words (nouns, verbs, adjectives, pronouns, function words, particles).

For some categories characteristic developmental patterns can be observed. For nouns there is an obvious increase until age 1;9 followed by a slowdown. Verbs grow continuously in a linear fashion. The linear increase of verbs combined with the observation that verbs gradually 'overtake' nouns resembles the pattern of verb development found in Bassano (2000) for French. The proportion of relational words decreases abruptly at first, then more slowly. The late but strong increase in function words points to a developmental step which may be connected to a transition from the one-word stage to a stage where we see the beginning of combinatorial speech and sentence formation. A turn in development is also indicated by the use of onomatopoeic words, which are only present to a slight extent in the beginning, when social and personal-social words are the predominant categories. During the second year the proportion of onomatopoeic words, which children use to imitate the sounds of animals and objects and to accompany activities, gains ground. By 3;0, children's onomatopoeic words have largely been pre-empted by conventional lexical items of their target language.

CONCLUSION

In sum, the results reported here on the composition of the lexicon illustrate the dynamic development for lexical categories during early vocabulary acquisition. These results do not support a strong version of the noun-bias hypothesis because not nouns, but relational words and personal-social words dominate during the early stages. Moreover, there is no single point in time where the use of nouns outweighs other word categories. Nevertheless, the German data confirm the assumption that nouns appear prior to verbs in the developmental sequence. It is only in this sense that a weaker version of the noun-bias hypothesis could be upheld.

The pattern of lexical acquisition observed for German children indicates that early word use does not primarily serve to refer to objects. The early words used in interaction by the majority of the children, are better characterized as

(a) deictic expressions for referring to contents of shared attention
(b) relational words, which are used to ask someone to perform an activity (with or without an object)
(c) personal-social words, with which children regulate interactions or communicate their attitude toward ongoing events.

In view of these findings based on actual word usage we conclude that the hypothesis that in acquiring their vocabulary children predominantly rely on nouns must be seriously questioned, at least in its strong version. Further

insight into lexical development can be gained by addressing additional issues. For example, the patterns of vocabulary growth reported here need to be verified by examining shorter time intervals during the second and third years of life. The investigation of productive word use should be complemented by studies on children's receptive vocabulary, including explicit word-comprehension tests. With respect to future research on vocabulary composition, the impact of input on children's lexical development needs to be determined. Since data on the composition of adult speakers' vocabularies in German is not available at present, a detailed investigation of the distribution of word categories in German child-directed speech is clearly required.

REFERENCES

Anisfeld, M., Gasparini, D., Hoberman, M. J. & Rosenberg, E. S. (1998). Lexical acceleration coincides with the onset of combinatorial speech. *First Language* **18**, 165–84.

Augst, G. (1984). *Kinderwort : der aktive Kinderwortschatz* (*kurz vor der Einschulung*) *nach Sachgebieten geordnet mit einem alphabetischen Register*. Frankfurt/M. : Lang.

Bassano, D. (2000). Early development of nouns and verbs in French : exploring the interface between lexicon and grammar. *Journal of Child Language* **27**, 512–59.

Bates, E., Bretherton, I. & Snyder, L. (1988). *From first words to grammar. Individual differences and dissociable mechanisms*. Cambridge : CUP.

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**, 85–121.

Bates, E., Dale, P. S. & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (eds), *The Handbook of Child Language*. Oxford : Blackwell.

Behrens, H. (1998). How difficult are complex verbs ? Evidence from German, Dutch, and English. In E. V. M. Lieven (ed.), special issue : developing a verb category : cross-linguistic perspectives. *Linguistics* **36**, 679–713.

Bloom, L. (1993). *The transition from infancy to language : acquiring the power of expression*. New York : Cambridge University Press.

Bloom, L., Tinker, E. & Margulis, C. (1993). The words children learn : evidence against a noun bias in early vocabularies. *Cognitive Development* **8**, 431–50.

Bornstein, M. H. & Haynes, O. M. (1998). Vocabulary competence in early childhood : measurement, latent construct, and predictive validity. *Child Development* **69**, 654–71.

Brown, P. (1998). Children's first verbs in Tzeltal : evidence for an early verb category. In E. V. M. Lieven (ed.), special issue : developing a verb category : cross-linguistic perspectives. *Linguistics* **36**, 713–55.

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

Choi, S. (1998). Verbs in early lexical and syntactic development in Korean. In E. V. M. Lieven (ed.), special issue : developing a verb category : cross-linguistic perspectives. *Linguistics* **36**, 755–81.

Choi, S. (2000). Caregiver input in English and Korean : use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language* **27**, 69–96.

Clark, E. (1993). *The lexicon in acquisition*. Cambridge : CUP.

Dromi, E. (1987). *Early lexical development*. Cambridge : CUP.

Dromi, E. (1999). Early lexical development. In M. Barrett (ed.), *The development of language*. Hove : Psychology Press.

Elsen, H. (1999). Auswirkungen des Lautsystems auf den Erwerb des Lexikons – Eine

funktionalistisch-kognitive Perspektive. In J. Meibauer & M. Rothweiler (eds), *Das Lexikon im Spracherwerb*. Tübingen, Basel: Francke.
Fenson, L., Dale, P., Reznick, S., Bates, E., Thal, D. & Pethick, S. D. (1994). *Variability in early communicative development*. (Monographs of the Society for Research in Child Development **59**.) Chicago: The University of Chicago Press.
Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* **32**, 675–701.
Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory* **4**, 161–78.
Gentner, D. (1982). Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In S. Kuczaj (ed.), *Language Development, Vol. 2, Language, thought and culture*. Hillsdale, NJ: Erlbaum.
Goldfield, B. & Reznick, S. (1990). Early lexical acquisition: rate, content and the vocabulary spurt. *Journal of Child Language* **17**, 171–81.
Goldfield, B. & Reznick, S. (1996). Measuring the vocabulary spurt: a reply to Mervis & Bertrand. *Journal of Child Language* **23**, 241–46.
Gopnik, A. (1988). Three types of early words: the emergence of social words, names and cognitive-relational words in the one-word stage and their relation to cognitive development. *First Language* **8**, 49–70.
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.
Kauschke, C. (1999). Früher Wortschatzerwerb im Deutschen: eine empirische Studie zum Entwicklungsverlauf und zur Komposition des kindlichen Lexikons. In: J. Meibauer & M. Rothweiler (eds), *Das Lexikon im Spracherwerb*. Tübingen: Francke.
Kauschke, C. (2000). *Der Erwerb des frühkindlichen Lexikons – eine empirische Studie zur Entwicklung des Wortschatzes im Deutschen*. Tübingen: Narr.
Kim, M., McGregor, K. K. & Thompson, C. K. (2000). Early lexical development in English- and Korean-speaking children: language-general and language-specific patterns. *Journal of Child Language* **27**, 225–54.
Klann-Delius, G. (1990). Transkriptionsmanual. Unpublished manuscript. Berlin.
Klann-Delius, G. (1992). *Manual zur Kodierung dialogischer Austauschprozesse in der Mutter-Kind-Interaktion*. Unpublished manuscript. Berlin.
Klann-Delius, G., Hédervári, E. & Hofmeister, C. (1996). *Die Bedeutung der emotionalen Qualität der Mutter-Kind-Interaktion für den Erwerb der Dialogfähigkeit des Kindes: Abschlußbericht*. Unpublished manuscript. Berlin.
Lieven, E. V. M., Pine, J. & Dresner Barnes, H. (1992). Individual differences in early vocabulary development: redefining the referential–expressive distinction. *Journal of Child Language* **19**, 287–310.
Menyuk, P., Liebergott, J. W. & Schultz, M. C. (1995). *Early language development in full-term and premature infants*. Hillsdale, NJ: Erlbaum.
Mervis, C. & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: a response to Goldfield & Reznick. *Journal of Child Language* **22**, 461–68.
Nelson, K. (1973). *Structure and strategy in learning to talk*. (Monographs of the society for research in child development, **38**, 1–2.) Chicago: University Press.
Nelson, K., Hampson, J. & Kessler Shaw, L. (1993). Nouns in early lexicons: evidence, explanations and implications. *Journal of Child Language* **20**, 61–84.
Pine, J. M., Lieven, E. V. M. & Rowland, C. F. (1996). Observational and checklist measures of vocabulary composition: what do they mean? *Journal of Child Language* **23**, 573–89.
Pine, J. M., Lieven, E. V. M. & Rowland, C. F. (1997). Stylistic variation at the 'single-word' stage: relations between maternal speech characteristics and children's vocabulary composition and usage. *Child Development* **68**, 807–819.
Pregel, D. & Rickheit, G. (1987). *Der Wortschatz im Grundschulalter. Häufigkeitswörterbuch zum verbalen, substantivischen und adjektivischen Wortgebrauch*. Hildesheim, Zürich. New York: Olms.

Robinson, B. F. & Mervis, C. B. (1998). Disentangling early language development: modelling lexical and grammatical acquisition using an extension of case-study methodology. *Developmental Psychology* **34**, 363–75.
Smith, C. & Sachs, J. (1990). Cognition and the verb lexicon in early lexical development. *Applied Psycholinguistics* **11**, 409–24.
Snyder, L., Bates, E. & Bretherton, I. (1981). Content and context in early lexical development. *Journal of Child Language* **8**, 565–82.
Stern, C. & Stern, W. (1928). *Die Kindersprache. Eine psychologische und sprachtheoretische Untersuchung*. (4th edition, re-print 1987). Darmstadt: Wissenschaftliche Buchgesellschaft.
Tardif, T. (1996). Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. *Developmental Psychology* **32**, 492–504.
Tardif, T., Shatz, M. & Naigles, L. (1997). Caregiver speech and children's use of nouns and verbs: a comparison of English, Italian and Mandarin. *Journal of Child Language* **24**, 535–65.
Tracy, R. (1991). *Sprachliche Strukturentwicklung. Linguistische und kognitionspsychologische Aspekte einer Theorie des Erstspracherwerbs*. Tübingen: Narr.
Vihman, M. & McCune, L. (1994). When is a word a word? *Journal of Child Language* **21**, 517–42.
Whong-Barr, M. K. (2000). The noun bias contested: implications for syntactic bootstrapping. *Newcastle & Durham Working Papers in Linguistics*, Vol. 6, 159–167.