Introduction

The acquisition of the lexicon is a central and complex component of child language development. It has received growing interest within psycholinguistic research and is becoming a field of study in its own right. At the same time, lexical development also interacts with acquisition processes in other linguistic domains. Therefore, early lexical capacities might be indicative of subsequent language skills. It is a question of crucial importance, whether certain capacities belong to the normal range of individual variation, or whether these capacities should rather be considered as an indication of a delayed or disturbed language acquisition process. To answer this question, empirically based knowledge about the normal course of vocabulary development within a particular language has first to be established. In this paper, a longitudinal study on early lexical acquisition in German is presented which deals with both descriptive and predictive aspects of lexical development.

1. Lexical development

1.1 Vocabulary growth

Concerning the course of lexical development in undisturbed circumstances, a lot of empirical studies have been carried out, focusing on two major topics: on the one hand, patterns of lexical growth as a function of age have been examined. On the other hand, the composition of the infant lexicon has become a controversial subject of research. The following sections present a brief outline of some relevant findings in these domains.

The emergence of the first words constitutes an important milestone in infant language acquisition. During this initial phase, beginning around the first birthday, vocabulary growth is typically very slow. The rate at which children continue to accumulate new words in the second and third year of life differs individually. In looking at single cases as well as larger samples, varying growth patterns have been demonstrated. In many cases a sudden acceleration in the rate of vocabulary growth has been reported around the age of 18-19 months. This widespread phenomenon of a rapid and sudden growth is referred to as vocabulary spurt (for a summary, see Dromi 1999). Whether the vocabulary spurt is a hallmark in the development of every child is unclear, as the controversy between Goldfield & Reznick (1990, 1996) and Mervis & Bertrand (1995) reveals. Other patterns are a gradual and linear increase of the vocabulary, an alternating pattern of spurt intervals and plateaus or an extended spurt phase. It cannot be excluded that differences in the empirical findings are due to methodological factors including different methods of sampling (checklists versus observational measures) and differences in the time span of data collection and the time intervals between sampling points. Although there is a great deal of individual variation in children's lexical
development, a group of children has been identified in recent studies, which is characterised by a late onset and a slow progress of expressive lexical development. I will return to the language development profiles of the so-called Late Talkers in section 2.

1.2. Vocabulary composition

Turning to the issue of vocabulary composition, there has been a constant debate on the type of word categories present and/or dominant in the lexicon of infants. Alongside the hypothesis that children initially prefer nouns and refer predominantly to objects, we find proponents of the view that the composition of early vocabularies is more heterogeneous. Gentner (1981, 1982) formulated the so-called noun bias-hypothesis, which states that nouns are acquired early and easily, because they are formally and conceptually less complex than other word categories (especially compared to verbs). The dominance and early appearance of nouns was said to be valid across different languages. However, recent crosslinguistic studies challenge the notion of a universal noun bias. 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. In the lexical development of children acquiring Mandarin and Korean (e.g., Choi 1998, Tardif 1996, Tardif et al. 1997), 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 and frequent use of verbs in these languages is linked to features of the input, such as frequency and saliency of verbs, or to the pragmatic focus of the caregiver's utterances. In English, nouns actually emerge prior to 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 (see Gopnik 1988, Dromi 1999). Bates et al. (1994, 1995) conducted an extensive cross-sectional study in order to determine the time of appearance and the proportion of different word classes in English. While the overall vocabulary grew constantly, the different word categories made up varying proportions at different points in time. Three 'waves' of internal reorganization were found. 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. The results of some studies (e.g., Bloom et al. 1993) suggest that nouns amount to one third at the "50 words"-stage. Thus, not even for English do nouns constitute an overwhelming quantitative majority.

Taken together, it can be concluded that the role played by nouns during vocabulary development does not seem to be as prominent as has been assumed. Moreover, language-specific aspects have to be carefully considered.

2. Developmental associations in language acquisition

The next point to be considered is the relation of lexical development to other domains in language acquisition. Correlations between developmental steps in the domains of phonology, lexicon, and syntax at specific points in time have been found in various studies of normally developing children (Bates et al. 1988, Bates et al. 1994, 1995, Fenson et al. 1994, Menyuk et al. 1995, Tamis-LeMonda 1998, Thal et al. 1995). For instance, research has demonstrated that early lexical abilities are a predictor of
later performance in the areas of morphological and syntactic complexity. The emergence and elaboration of grammar seems to be highly dependent on vocabulary size (for a review, see Bates & Goodman 1999). It is reasonable to infer from this that limitations in the early lexicon will lead to limitations in other domains. This assumption is confirmed by investigations of Late Talkers, children whose linguistic development is characterised by delayed and slow lexical development. Results from relevant studies show that reduced vocabulary size predicts a delay in grammatical development (Bishop & Edmundson 1987, Rescorla & Schwartz 1990, Paul 1991, Scarborough & Dobrich 1990, Ellis Weismer et al. 1994, Whitehurst & Fischel 1994, Bates et al. 1997, Rescorla et al. 1997, among others). Thus, grammatical development appears to be tied to lexical level (Bates & Goodman 1999: 54). Some of the late talking children, the so-called Late Bloomers, will attain normal linguistic abilities after catching up with respect to vocabulary. In other cases, deficits in the morphological and syntactic domain persist, even as vocabulary grows. Children with this type of development are said to display Specific Language Impairment.

At the opposite end of the normal distribution we find children who show early and quick lexical development paired with early or above-average development of grammatical abilities (Early Talkers, Thal et al. 1996, 1997). To conclude, developmental associations between lexical and grammatical development become evident for the population of normally developing children as well as for Late and Early Talkers.

3. A study on lexical development in German

3.1. Subjects and methods

Most of the findings reported so far have been limited to the English language, so that any conclusions on early lexical acquisition in German are premature due to the lack of systematic empirical studies. For that reason, the longitudinal study reported below aims to document lexical development in German children in their second and third year of life and to evaluate individual differences and their consequences for further language development.

The corpus consists of spontaneous speech samples from 32 children (16 boys, 16 girls) at the ages of 13, 15, 21 and 36 months. The children were videotaped in a free play situation with their mothers. Based on 10-minute transcripts of spontaneous speech production, the following properties of the collected data were analysed:

- size of vocabulary and frequency of word use in relation to age (development of types and tokens)
- proportion of word categories
- individual differences in vocabulary size
- syntactic and morphological abilities at the age of three and their relation to earlier lexical abilities.

For details on the methods applied for transcription, coding of word categories, analysis of grammatical abilities and application of statistics, see Kauschke (1999, 2000) and Kauschke & Hofmeister (submitted).

The results on vocabulary growth and composition will be presented first, followed by findings on the relation between early lexical and later grammatical performance. Finally, two single cases are chosen to illustrate the group results.
3.2. Results on vocabulary growth

Not surprisingly, the use of both types and tokens increases in the course of lexical development (see Table 1):

<table>
<thead>
<tr>
<th>age</th>
<th>13 months</th>
<th>15 months</th>
<th>21 months</th>
<th>36 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>total number of</td>
<td>82</td>
<td>187</td>
<td>649</td>
<td>2522</td>
</tr>
<tr>
<td>types</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean types</td>
<td>2.56</td>
<td>5.84</td>
<td>20.28</td>
<td>78.81</td>
</tr>
<tr>
<td>range types</td>
<td>0-10</td>
<td>0-16</td>
<td>0-54</td>
<td>10-138</td>
</tr>
<tr>
<td>SD types</td>
<td>2.18</td>
<td>3.8</td>
<td>12.17</td>
<td>31.79</td>
</tr>
<tr>
<td>total number of</td>
<td>302</td>
<td>556</td>
<td>1846</td>
<td>6411</td>
</tr>
<tr>
<td>tokens</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean tokens</td>
<td>9.44</td>
<td>17.38</td>
<td>57.69</td>
<td>200.34</td>
</tr>
<tr>
<td>range tokens</td>
<td>0-52</td>
<td>0-58</td>
<td>0-140</td>
<td>18-485</td>
</tr>
<tr>
<td>SD tokens</td>
<td>10.66</td>
<td>12.89</td>
<td>36.00</td>
<td>104.18</td>
</tr>
<tr>
<td>type-token-ratio</td>
<td>0.44</td>
<td>0.40</td>
<td>0.38</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Table 1: descriptive values for types and tokens

The growth rate for the number of different words (types) constitutes a significant increase as a function of age. Age accounts for 84% of the variance in the number of types. The frequency of word use (tokens) also increases significantly with age. So, general developmental sequences in lexical production become evident.

The number of types as well as the number of tokens increases in a non-linear fashion from age 1;1 to 1;9 and levels out at the final sampling point. Trend analyses of the logarithms of the type and token values indicate an exponential component during the second year with a subsequent slow-down of the growth rate. Figures 1a and 1b show the growth curves of the types and tokens, illustrating a vocabulary spurt-like pattern in the second year, observable in both the number of different words and the total number of words produced.
Additionally, there is a clear correlation between types and tokens, where the type-token-ratio remains constant. An increase of types is always accompanied by an increase of tokens. Children with limited vocabularies do not tend to use their words more often.

### 3.2. Results on vocabulary composition
The vocabulary composition findings illustrate the dynamics in the development of word categories. In the beginning, relational words (e.g., *da* "there", *weg* – "gone"), personal-social words (e.g., *ja* "yes", *tschüß* "bye") are predominant. At the first sampling point, these two categories constitute more than three-quarters of the lexicon. From the outset, nouns are present and their proportion grows in particular during the second year of life. At the age of three years the proportion of nouns is below 25% in all children. At the age of 15 months verbs begin to appear. Their proportion increases considerably and gradually until they constitute the largest proportion at 36 months. Adjectives are also present in the lexicon from the outset. Their proportion stays relatively small in the course of development. At the later sampling points, function words begin to emerge and they increase over time.

Trend analyses (on the proportions of types) clarify characteristic developmental patterns of certain word categories (see figure 2 for a selection of word categories). For verbs, adjectives, pronouns, function words and the category of other words (comprising mainly particles) linear increasing trends were found. The proportions of personal-social words and relational words show linear decreasing trends. Additionally, the proportion of nouns can be described as a significant non-linear trend. There is an early increase of nouns with a peak frequency of 25% at 21 months, followed by a slowdown in the third year.

In sum, the categories dominating in the initial phase of lexical development (relational and personal-social words) are gradually complemented with nouns, verbs, function words and other word classes. This process leads to a balanced lexical composition by the age of three. At that age, the composition of the type set does not differ considerably from the composition of the token set.
3.3. Individual differences in lexical size and their consequences for grammatical complexity at age three

During the second year, strong individual differences in vocabulary size are observed. The number of different words and the total number of words increasingly vary across children (see Table 1: range and standard deviations). These individual differences cannot be attributed to gender, socio-economic status or the quality of attachment between mother and child. Rank order tests show significant differences between the children, which remain stable over an extended period of time (i.e., from 13 to 36 months). Some children constantly belong to the upper end, while other children remain in the lower range. The longitudinal consistency is supported by significant correlations between the word production values per child at the four sampling points (for types and tokens). Thus, individual differences in the number of words produced during the second year seem to be a useful indicator for subsequent lexical development in the first place. Children with limited lexical production in the second year of life will not overcome this disadvantage; they continue to rely on a restricted vocabulary at the age of three. These results support a developmental continuity within the lexical domain.

The next step was to determine whether quantitative differences in early lexical development are related to other language capacities at the age of three. Various
morphological and syntactic aspects of the children's spontaneous speech at 36 months were examined.

The variables were, for example, MLU and Upper Bound, complexity of utterances, appearance of subordinate clauses, omissions of obligatory elements or the capacities in verb inflection and case marking (for an overview of the syntactic and morphological abilities of the 32 subjects see Table 2). Correlation measures, factor analyses and regression analyses were performed. Taken together, the results indicate highly significant relations between vocabulary production in the second year and later grammatical capacities.

<table>
<thead>
<tr>
<th>variable</th>
<th>mean (range)</th>
<th>standard-deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>utterance length</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLU in words</td>
<td>3 (1-3,97)</td>
<td>0,64</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>6,81 (1-12)</td>
<td>2,33</td>
</tr>
<tr>
<td><strong>analysed utterances:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion of multiword utterances</td>
<td>57%</td>
<td>0,18</td>
</tr>
<tr>
<td>proportion of utterances with multiple constituents</td>
<td>47%</td>
<td>0,19</td>
</tr>
<tr>
<td>proportion of subordinate clauses</td>
<td>1%</td>
<td>0,02</td>
</tr>
<tr>
<td>proportion of omissions</td>
<td>31%</td>
<td>0,19</td>
</tr>
<tr>
<td>proportion of false or reduced sentences</td>
<td>25%</td>
<td>0,16</td>
</tr>
<tr>
<td><strong>utterances including verbs:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion of correct verb inflections</td>
<td>86%</td>
<td>0,11</td>
</tr>
<tr>
<td>proportion of false verb inflection</td>
<td>2%</td>
<td>0,11</td>
</tr>
<tr>
<td>proportion of verbs without inflections</td>
<td>12%</td>
<td>0,11</td>
</tr>
</tbody>
</table>

Table 2: means and standard deviations of some grammatical variables

In more detail, several single variables from the grammatical analysis, among them MLU, show a high correlation with the number of types produced earlier. These relations appear to be stable at 21 months (see Table 3). Some variables (e.g., MLU) already correlate with the type values at 13 months. The variables indicating grammatical problems (e.g., proportion of omissions or uninflected verbs) correlate negatively with the type values.
Table 3: regression analyses

Factor analyses were performed in order to structure and reduce the data. First, the type values at 13, 15, and 21 months were combined into one factor. The second factor contained several grammatical variables, which had proved to be relevant in preceding factor analyses. Individual scores for every child were extracted. A regression analysis indicates a highly significant relationship between the predictor variable (lexical size) and the dependent variable (grammar): p = 0.0006. Vocabulary size in the second year explains one-third (34%) of the individual variation in the later grammatical scores. Outlier's tests do not show a distortion of the results due to extreme single values. These results suggest an associated and successive development of lexical and grammatical processes.

The scatterplot in Figure 3 illustrates a non-linear relationship between the two factors, supporting the notion that some "critical mass" of lexical items is required for an age-appropriate acquisition of grammar.
3.3. Single cases

The group results can be illustrated by the language profiles of single cases. Two girls – one at the upper and one at the lower end of the distribution – demonstrate an advanced versus an impaired course of both lexical and grammatical development.

The Early Talker produces many more words than the other children, and word combinations are attested very early (two-word combinations already appear at the age of 13 months). The girl's lexicon increases - with a quantitative spurt during her third year -, and continuously becomes differentiated. A dominance of nouns in the second year is followed by an expansion of verbs. In the meantime, her syntactic and morphological abilities become increasingly more refined. At the age of three, she has a productive vocabulary of substantial size, a rich verb lexicon, and her speech exhibits evidence of advanced grammatical phenomena.

In contrast, the Late Talker provides only a rudimentary expansion of her lexicon. Very few words are produced during the second and third year, none of which are verbs. In addition, there are almost no signs of grammatical development; the spontaneous speech does not exceed the one-word stage. Given the reduced and undifferentiated vocabulary, the complete lack of verbs, and the stagnation in the domains of syntax and morphology, the girl's language development at the age of three shows a sharp lag in comparison to the other children. It does not seem appropriate, in the light of the clear deviance from the mean group values, to judge her as "nothing more than the lower end of the normal curve of language development" (Whitehurst & Fischel 1994). Instead, this child stands out right from the outset and gradually develops a Specific Language Impairment. Even at the age of eight, persisting deficits are reported in her language-specific capacities as well as in her communicative and social behaviour. Thus, this is a
clear-cut case of negative long-term development of a child who showed an initial lexical delay. This kind of development with spreading and cumulative deficits is frequently reported in Late Talker studies.

Examining individual profiles among all children investigated in this study, the pattern of a roughly sufficient vocabulary production in the second year, combined with later problems in the domains of syntax and morphology, could also be demonstrated. However, there were no children with limited early vocabularies, combined with advanced grammatical skills at the age of three. That means that a good outcome in the presence of early lexical limitations is unlikely, but problems in the domains of syntax and morphology can also occur without earlier lexical deficits. These observations support the idea that a satisfactory level of lexical development is a prerequisite for grammatical development.

4. Summary and conclusions

Empirically supported claims regarding the development of children's vocabulary have so far been made mainly about the English language. The results presented here on spontaneous use of words in 32 German children contribute to a more differentiated picture of the process of lexical development in German children in the early stages.

The observed exponential increase in spontaneous word production in the second year, followed by a further deceleration, may best be described as vocabulary growth within a 'region of acceleration' (Bates et al. 1995: 105). Although the findings clearly support a general trend of a vocabulary spurt phase in the second year, there are individual patterns that differ from the group results in showing plateaus or linear phases of vocabulary growth or a spurt in the third year (e.g., the Early Talker introduced above), supporting the view that "there may be more than one way to build a lexicon" (Goldfield & Reznick 1990: 182).

The vocabulary composition findings show clearly 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, while the categories that are initially dominant undergo a relative decline. For some categories characteristic developmental patterns were observed (i.e., an early increase of nouns followed by an ensuing slowdown, a continuous, linear growth of verbs and a late increase of function words), comparable to the results of Bates et al. (1994) and Pine et al. (1997). The spontaneous speech data for German do not support a strong noun bias hypothesis, because of two reasons: nouns are not the first word category to emerge, and there is no single point of time where the use of nouns outweighs other word categories. However, a temporal asynchrony in the emergence of nouns and verbs is confirmed for German.

The evident continuity between lexical and grammatical development strongly suggests that developmental processes during early language acquisition are associated with each other. It seems thus reasonable to consider early lexical abilities an indicator for later grammatical complexity. Accordingly, the study shows that lexical limitations successfully serve as a reliable, early predictor of potential language acquisition problems and sometimes, even of severe and continuing disorders. This evaluation supports the results of other studies (see Rescorla et al. 1997) in which lexical development is taken to be a valid predictor of further language acquisition. Furthermore, it seems reasonable to conclude from the results that a satisfying level of
lexical competence is a necessary, but not sufficient condition for the emergence of productive grammar.

Endnotes

1 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.

References


