A normative study of early language development in German using an adaptation of the long and short version toddler CDI

Gisela Szagun [1, 2], Barbara Stumper [3], Satyam A. Schramm [2]

[1] University College London, UK
[2] University of Oldenburg, Germany
[3] Max Planck Institute for Evolutionary Anthropology Leipzig, Germany

Note:
This manuscript was submitted to *Journal of Child Language* in 2011. It reived very positive reviews from two reviewers and the Editor, but unfortunately could not published in JCL because we had already published the study in our German edition of FRAKIS, the name of our German CDI.

Address for correspondence:
Dr. Gisela Szagun
Developmental Science Research Department
University College London
2 Wakefield Street
London WC1N 1PF
UK
Email: gisela.szagun@googlemail.com

Abstract

The present paper reports norming data for German-speaking children's early language development using an adaptation of the CDI parental questionnaire. A long form and short form version were developed. For the long form questionnaires were collected for 1240 children between 1;6 and 3;0. Two-hourly spontaneous speech samples were collected from 60 children. Results showed extensive variability in the development of vocabulary, inflectional morphology and sentence complexity, paralleling that observed in other languages. Socioeconomic class, gender, and birth order had significant effects on language development. Our questionnaire demonstrates very good reliability and concurrent validity, the latter being tested using measures derived from spontaneous speech. Results for the short form version showed high concurrent validity for the short word list. Simulated scores based on the large sample and validated empirically on 411 participants are used as norm data. The extent of variability in vocabulary growth parallels that of the long vocabulary list.

Introduction

Variability in early language development has been documented impressively with the large-scale studies of American English using the *MacArthur-Bates Communicative Development Inventories* (CDIs) (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994; Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007, 2nd edition), a parent report instrument. In the acquisition of German variability has been less of a focus. While a number of studies offer important insights into the acquisition of vocabulary and the course and mechanisms of the acquisition of many morphological paradigms of German (Mills, 1985; Clahsen, 1984; Behrens, 1993, 2002; Szagun, 2001, 2004a; Kauschke & Hofmeister, 2002; Szagun, Stumper, Sondag & Franik, 2007; Abbot-Smith & Behrens, 2006; Brandt, Diessel & Tomasello, 2008), the available corpora in CHILDES (MacWhinney, 2009) are not large enough for assessing the variability in the population and for obtaining age norms for early vocabulary and grammatical development.

Evidence for variability in early language development based on representative samples challenges the view that in acquiring language all normal children go through the same stages at the same ages (Stromswold, 2000). The first such evidence came from a study on the acquisition of American English using parent report data obtained with the Communicative Development Inventories (CDI) (Fenson et al., 1994). Since then normative studies using the same methodology have shown that there is extensive variability in the growth of early vocabulary and grammar in a variety of languages (*Italian*: Caselli, Casadio & Bates, 1999; *Swedish*: Berglund & Eriksson, 2000; *British English*: Hamilton, Plunkett & Schafer, 2000; *Dutch*: Zink & Lejaegere, 2002; *European Spanish*: López Ornat, Gallego, Gallo, Karousou, Mariscal & Martínez, 2005; *French*: Kern, 2007; *Danish*: Bleses, Vach, Slott, Wehberg, Thomson, Madsen & Basbøll, 2008a, b). Normative data from German-speaking children would add another language. With more and more evidence from different languages it may turn out that what is universal in language development is not a fixed timetable of linguistic achievements but extensive variation between children.

From a clinical viewpoint knowledge about the range of variability to be expected in early language development and clear criteria for identifying what may constitute delay are extremely important. The lack of age norms for the early development of German has posed significant problems for health care and educational professionals. In Germany, language assessment is part of paediatric preventive health checks for two-year-olds. Currently, there is uncertainty about when to classify a young child as ‘outside the norm’. A variety of language assessment instruments exist (Grimm & Doil, 2000; von Suchodoletz, 2008; Bockmann & Kiese-Himmel, 2008), but none are based on population norms. When these instruments are used, the proportion of young children characterised as ‘language delayed’ or even ‘language impaired’ varies between 15 % and 28 % depending on which instrument is used. Perhaps the most influential one is the parental questionnaire ELFRA (Grimm & Doil, 2006) which is based on *The Language Development Survey* (Rescorla, 1989) and the CDI (Fenson et al., 1994). The purpose of this instrument is to identify children at risk of language impairment. A

child is classified ‘at risk’ if his/her vocabulary is smaller than 50 words out of the instrument’s short vocabulary checklist of 260 words at the age of 2;0. According to this criterion 20 % of German-speaking two-year-olds are at risk of developing language impairment (Grimm & Doil, 2006). This proportion is likely to be inflated, as the 50-word criterion was set following Rescorla (1989) who used it for American English which is a language with particularly fast early vocabulary growth (Bleses et al., 2008a). Criteria for delay which are based on age norms and the extent of variability typically observed in the population of one- to three-year old German-speaking children are likely to be more appropriate.

Parent report instruments enable researchers to obtain normative data on early child language development as they allow data collection from large samples. The perhaps most widely known ones are the *MacArthur-Bates Communicative Development Inventories* (CDI) (Fenson et al., 1994; Fenson et al., 2007). Having initially been developed for American English they have by now been adapted to many different languages (for an overview see Bleses et al., 2008) and have proven to be highly valid and reliable instruments for the assessment of early child language. We began with the adaption of the CDI toddler inventory to German and reported first results with 333 participants earlier (Szagun, Steinbrink, Franik & Stumper, 2006). The results of this initial study revealed extensive variability in the growth of vocabulary and grammar in German-speaking children, a strong association between vocabulary and grammatical growth and dependence of grammatical development on vocabulary size (Szagun et al., 2006). In these respects the results for German resemble those in other languages (Bleses et al., 2008a). However, there are several limitations of this earlier study. One is its sample size. It allowed only the use of the standard deviation as an index of variability and no fine-grained analysis of the distribution of children per age group in terms of percentiles. The number of children per monthly age group was too small to establish age norms. Another limitation is the lack of demographic information regarding socioeconomic status and children’s birth order. These limitations restrict comparisons with norm data from other cultures, the investigation of possible sources of variability and the use of the data as a basis for criteria of delay in development.

The present study aims at obtaining normative data on the typical course and extent of variability in vocabulary and grammatical development in German-speaking toddlers between the ages of 1;6 and 2;6. Additionally, it investigates the influence of the demographic factors of gender, birth order and socioeconomic status on such development. The questionnaire used is very similar to the one in the earlier study (Szagun et al., 2006). However, there are some minor alterations in the vocabulary and grammar checklists, and there is the extension of the section on basic information including information on birth order and socioeconomic status. In keeping with the underlying concept of the CDIs in many languages the instrument assesses culture specific linguistic behaviour in the toddler period. It is designed to capture the transition from words to grammar which, typically, occurs between 1;6 and 2;6. The general structure of our CDI adaptation to German is in keeping with the structure of the original American CDI (Fenson et al., 1994; Fenson et al., 2007). It contains a vocabulary checklist organised in the same semantic fields, a sentence complexity scale designed to capture morphosyntactic changes indicative of increased sentence length and complexity, and an inflectional morpheme scale designed to capture the inflectional paradigms of German which children in the targeted age range typically begin to acquire (for more details, see Methods

Section and Szagun et al. 2006; Szagun, Stumper & Schramm, 2009). Due to the nature of the language, the inflectional morpheme section is considerably more extensive than the corresponding section in American English.

Results regarding the effects of demographic factors differ cross-linguistically. Consistent and stable effects of gender on language development have been reported using a variety of measures, with girls outperforming boys (Maccoby, 1966; Bornstein, Hahn & Haynes, 2004). Studies with the CDI in different languages are not always consistent with these findings. While an effect of gender favouring girls was observed for American English, Dutch and Danish (Fenson et al., 1994; Zink & Lejaegere, 2002; Bleses et al., 2008b), no such effect was found for Hebrew and Swedish (Maital et al., 2000; Berglund & Eriksson, 2000). Where they are reported, effects of socioeconomic status and birth order tend to be small in magnitude (Fenson et al., 1994; Fenson et al., 2007; Zink & Lejaegere, 2002) and sometimes absent (Hamilton et al., 2000). An investigation of the effects of gender, birth order and socioeconomic status in our German norming study will broaden the cross-linguistic perspective.

On the basis of our previous study we expect the extent of variability in the early development of German to be very large and parallel that in other languages. We also expect effects of the demographic factors of socioeconomic status and the child's gender and birth order. In Germany the majority of children under the age of three years grow up in family environments with the mother as the main caretaker and the main source of language input. Such an environment is likely to be more beneficial for children of more highly educated mothers, as these offer rich language input which in turn mediates the influence of socioeconomic status (Hoff, 2003; Clark, 2003). It would also benefit first born children as their main language input is from the care-taking adult. Gender effects are seen to be due to an interaction of biological, psychological and social variables (Bornstein et al., 2004). A family environment with the mother as the main caretaker could place more importance on gender-role stereotyping and behaviour. For this reason and on the basis of our previous results with a smaller sample Szagun et al., 2006) we expect faster language growth in girls.

A further aim of the present study is to provide measures of reliability and validity for our parent report instrument. This concerns the internal consistency of the scales, their test-retest reliability and their concurrent validity. In our view, the most stringent method of assessing the validity of parent report is by relating it to children's spontaneous production of language in a naturalistic setting. This is why we designed our study to collect spontaneous speech samples from a subsample of children. The degree of agreement between measures of vocabulary and grammar based on parent report and on spontaneous speech from the same children can then be calculated.

In view of the increased demand for instruments to assess early child language within the short time available for a paediatric preventive health check or other clinical settings, a short form version of our parent report instrument was developed. Its development, norming, empirical validation and psychometric properties will be reported here. When presenting parents with a short vocabulary list, the selection of items for this list is crucial and it has to be ascertained that the items discriminate maximally between children with slow and fast language development. Such an instrument should be constructed on the basis of normative data about the typical course and extent of variability of language development. For German,

therefore, the first step consisted in obtaining normative data with the comprehensive parent report instrument.

Long Form Version

Method

Description of the questionnaire

The German parent report questionnaire contains a vocabulary part, a grammar part and a basic information sheet. The vocabulary and grammar parts use a recognition format. The vocabulary part is closely modelled on the CDI (Fenson et al., 1994). A checklist of 600 words in the 22 semantic fields used by Fenson et al. (1994) is presented and parents are asked to indicate whether their child produces a particular word. The 600 words were selected from a word list of all the words occurring in the corpora of 22 German-speaking children aged between 1;4 and 2;10 (Szagun, 2004a) comprising 170 two-hourly samples of spontaneous speech. All the words used in the German spontaneous speech data which were German equivalents of English words contained in the Fenson et al. (1994) CDI were included in the vocabulary list. These comprise 75% of words in the German list. Additionally, German equivalents of some words occurring in the American CDI but not in the German speech samples were included. This was done for words which denote objects and events in the lives of small children in both cultures but may not have been picked up because our recordings occurred in the more restricted environment of a University playroom and not at home (for details see Szagun, 2001; 2004b). These words amount to 11% of the German word list. Words which occurred in the American checklist but did not seem appropriate to German culture were not included. Most of these concerned items in the categories “food and drink”. Words which were culture-specific for German and occurred in our speech samples but not in the American list were added, e.g. *Autobahn* (motorway), *Fußball* (football). These were 14% of the word list. Thus, out of 600 words of the German vocabulary checklist 86 % are German equivalents of words in the American CDI (Fenson et al., 1994) and 14 % were words which do not have English equivalents in the original CDI. The German vocabulary list is kept shorter than the original CDI word list (Fenson et al., 1994) because the grammar sections are longer, and the aim is to keep the total questionnaire at a reasonable length. Most of the 22 semantic categories were shortened and, where possible, this was done by leaving out the words which were not culturally appropriate. However, the categories “quantifiers and articles” and “helping verbs” had to be slightly lengthened to include inflected forms. Vocabulary is scored by summing up the total of words produced.

The grammar part has two sections, one on inflectional morphology and one on sentence complexity. Due to the more highly inflected nature of German the section on inflectional morphology differs considerably from its equivalent in American English (Fenson et al., 1994). The section assesses five grammatical paradigms all of which are known to emerge in the age range from 1;6 to 2;6 (Mills, 1985; Behrens, 1993; 2001; Szagun, 2001a, b, 2004; Szagun et al., 2007). These are: 1) noun plurals, 2) gender marking on articles, 3) case marking on articles, 4) verb marking on main verbs for present tense and past participle, 5) modal verbs and forms of the copula. The section starts with four introductory questions asking parents to what extent children use inflectional morphemes giving examples of plurals,

articles, present tense and past participles. The “not yet”, “sometimes”, “often” answer format is used. Then a more detailed assessment of each of the five grammatical paradigms follows. For plurals examples of noun plurals in the different plural classes of German are presented. For gender marking examples of combinations of gender marked definite/indefinite article + noun and of adjective + noun are presented. The section on case marking presents examples of short utterances with case marked definite and indefinite articles. For marking of verbs on main verbs examples in the present tense singular and plural and past participles of different participle classes are presented. For modals and forms of the copula short utterances containing different modal verbs and forms of the copula are presented. All the items are drawn from the German spontaneous speech data (Szagun, 2004a). The parent is asked to indicate if the child produces an inflectional form or not either by marking a “yes” or “no” box or by circling the inflected form in an example utterance. A score is calculated per inflectional paradigm. The child is given one point per class of examples or per encircled example utterance. The score per paradigm consists of the summed points. Maximum scores differ per paradigm. They are: Plural = 6, gender = 8, case = 6, verbs = 10, modals and copula = 12. In addition to scores per paradigm, a total inflectional morpheme score is calculated consisting of the sum of the separate scores with a maximum value of 42. For more detail and examples of the different inflectional paradigms see Szagun et al. (2006).

The section on sentence complexity starts by asking parents if their children already produce word combinations, again using the “not yet”, “sometimes”, “often” answer format. Parents are then asked to write down the child’s three longest sentences they can remember. The section continues with a set of 32 blocks of sentences in pairs of different levels of complexity. This section is equivalent to the sentence complexity scale of the CDI (Fenson et al., 1994) in that it captures morphosyntactic changes which increase sentence length and complexity. For German this involves adding inflectional morphemes and, at a more advanced level, inflectional morphemes and function words such as auxiliaries, modals and articles. All example utterances are from naturalistic data (Szagun, 2004a). Parents are asked to tick the version of the sentence which is closest to what their child produces. One point is given, if the more complex version is chosen. The maximum score is 32.

The section on basic information contains questions regarding the child’s gender and date of birth, birth order, whether the child is a twin or was premature, there were serious medical problems, and whether the child is bilingual and, if so, which is the dominant language. Questions regarding parents enquire about the main caretaker’s educational status and who filled out the questionnaire.

Participants and data collection procedures

Parents of children between 1;6 and 2;6 were contacted for participation in the study via 15 paediatricians’ practices in different towns. The towns were Hannover, Oldenburg, Bremen, Essen (Ruhr), and a variety of smaller towns in northwest Germany. These towns represent densely populated large cities as well as medium and small sized towns and locations in Germany and are thus representative of the different environments people live in. In Oldenburg parents were also contacted via six day care centres. Additionally, parents from all over Germany who had heard about the study contacted us and took part. Most questionnaires were posted from the paediatricians’ practices without prior enquiry whether parents were willing to take part. Two practices distributed the questionnaires after prior contact and

agreement to participation. In the nurseries parents were asked by staff to take part in the study. Participation was voluntary. Parents were sent/given the questionnaire, a short letter asking for their cooperation and instructing them to complete the questionnaire within one day, a flyer with information about the study, and a consent form. Parent consent was required in writing. Parents were paid €5.00 for their participation.

The study is cross-sectional with one parent report collected for each child (except for a subsample of 60 children). The children are in 13 monthly age groups between ages 1;6 and 2;6. A subsample of 60 children, 20 at ages 1;6, 2;0 and 2;6 respectively, took part in the questionnaire study and a recording session of spontaneous speech. Per child a two-hour spontaneous speech sample was collected in a free play situation with a parent in a playroom at the University. The spontaneous speech samples were collected in order to test the external validity of our questionnaire measure with measures based on spontaneous speech. Everything spoken by the child was transcribed using CHILDES (MacWhinney, 2000). Parents in the subsample filled in the questionnaire twice within a 7- to 10-day span. These data served to assess the test-retest reliability. Parents in the subsample were paid €20.00.

Data collection took place between March 2005 and February 2007. Of the distributed questionnaires 37 % were returned. In 84% of the cases the questionnaires were filled in by the mother, in 16 % by mother and father. Of the 1425 questionnaires which were returned 185 (13 %) had to be discarded. Criteria for exclusion were: preterm birth, twin, serious medical problems, and German not as first language. Additionally, some children had to be excluded because the child was too young or too old or the questionnaire was insufficiently filled in. The final sample consisted of 1240 children.

Demographic characteristics of the sample

Table 1 shows the numbers of children in the 13 age groups for girls and boys separately and for all children together. In most age groups numbers of girls and boys are fairly even. Altogether there are 49 % girls and 51 % boys.

Table 1: Number of children in the age groups

Age in year and months	Girls	Boys	Children
1;6	42	45	87
1;7	40	40	80
1;8	42	57	99
1;9	40	40	80
1;10	53	44	97
1;11	49	50	96
2;0	53	56	109
2;1	41	67	108
2;2	45	51	96
2;3	45	47	92
2;4	47	58	105
2;5	51	40	91
2;6	57	40	97
Total	605	635	1240
in %	49 %	51 %	

As a measure of socioeconomic status we used mother's/father's educational level which is generally used as a measure of socioeconomic status in the CDIs (Fenson et al., 1994; Fenson et al., 2006; Zink & Lejaegere, 2002; Bleses et al., 2008b). Germany has a stratified school system, separating students according to their academic abilities. Mothers' educational level is measured here by type of school leaving certificate, adding University degree as a level. Table 2 shows the numbers and percentages of mothers in the different categories in comparison to the percentages for women between 20 and 40 in the population at large according to the population census figures of 2005 for the Federal Republic of Germany (Statistisches Bundesamt, 2006). The age span and gender correspond to the age and gender of main care takers of small children in Germany. As Table 2 indicates the proportion of women with high educational levels in our sample exceeds that in the population at large, and vice versa the proportion of women with a low educational level is lower. At the middle level there is good agreement.

Table 2. – Mothers' educational level in the present sample and in the total population

Level of schooling/education	Absolute number[a]	%	% according to 2005 census of women 20 to 40 years old
Hauptschule : Lowest, 9 years schooling	116	9.5	21.0
Realschule: Middle, 10 years schooling	421	34.5	32.1
Gymnasium: Highest, 13 years schooling	261	21.4	41.7[b]
University degree		424	34.7

[a] 18 cases were missing data.
[b] In the census this figure sums up persons with *Gymasium* (highest school level) and University degree.

Our sample includes bilingual children with German as first language. There are 97 (8 %) bilingual children in the sample. Twenty-three languages were named as second language. The most frequent ones were Polish and English, each accounting for 11 % out of the total of bilingual children, and Russian accounting for 10 %.

Results

Age-related trends of language growth

For the three language measures vocabulary, inflectional morphology and sentence complexity age-related trends are described and depicted in line graphs. We present children's scores at the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles per age group. This depicts the central tendency and the associated variation. Rather than raw values we are presenting fitted curves. This is done in order to compensate for fluctuations from month to month due to sampling (Fenson et al., 2007). In all cases the curve which gave the best fit to the data was the logistic

function. The line graphs presented here are for girls and boys combined (for percentiles for girls and boys separately, see Szagun et al., 2009).

Vocabulary

Figure 1 shows the growth of vocabulary between 1;6 and 2;6. In its central tendency, as indicated by the median, vocabulary increases from 41words at 1;6 to 486 words at 2;6. This is almost a 12-fold increase. Variability is considerable, especially between 1;9 and 2;3. At 2;0 the number of words of those 80 % children who lie above the 10th and below 90th percentile ranges between 48 and 456. The reduced difference between the median and the 90th percentile scores at 2;6 suggests a mild ceiling effect.

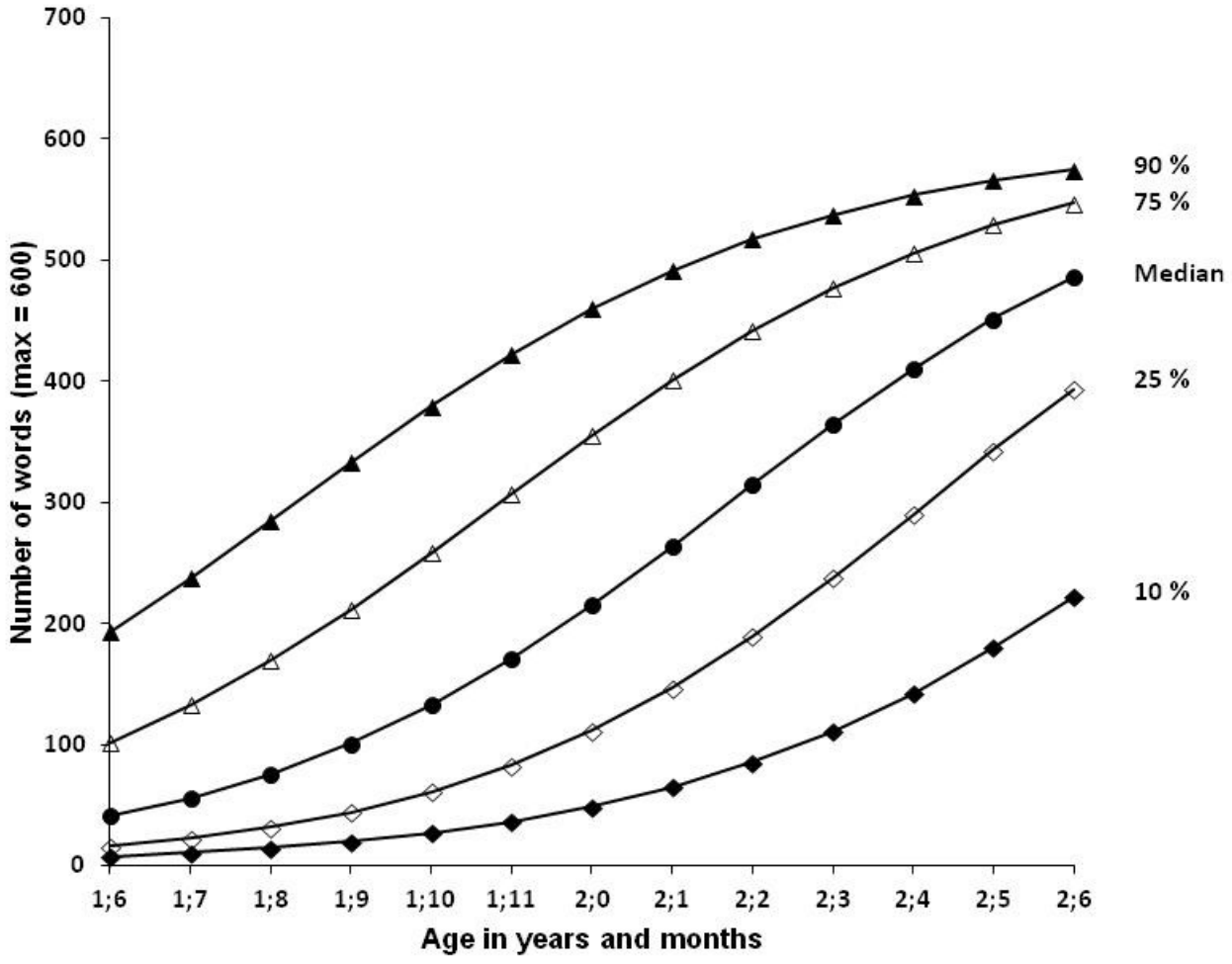

Figure 1. – Fitted percentile scores for words produced

Grammar

A number of measures were used to assess grammar (see *Methods Section*). Here, we will focus on results for the main grammar scales and subscales for inflectional paradigms. Results regarding the introductory questions on use of inflectional morphemes and word combinations are presented in Szagun et al. (2009).

Figure 2 shows the growth of inflectional morphology as measured by the inflectional morpheme scale which sums up the scores for the different inflectional paradigms. The maximum score is 42. As depicted in Fig. 2 the acquisition of inflectional morphology does not gain momentum until around 2;1 when considering its central tendency. As indicated by the median score increases from 0 at 1;6 to 30 at 2;6 with the largest increases between 2;1

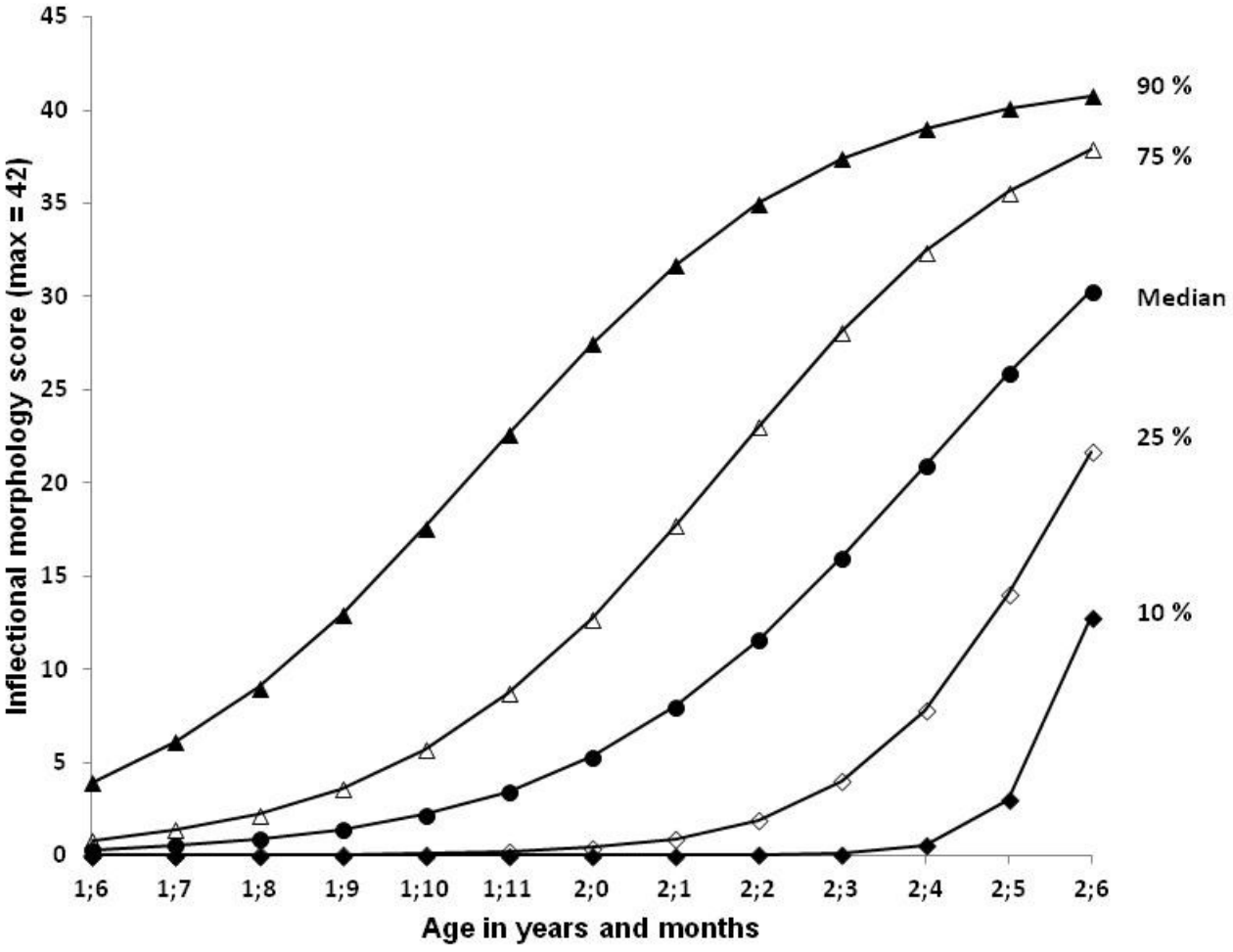

Figure 2. – Fitted percentile scores for inflectional morphemes produced

and 2;5. For the 90th percentile the largest increase is much earlier, between 1;9 and 2;2, and much later for the 10th percentile where growth does not begin until 2;4. Thus, there is a strong floor effect, but a milder ceiling effect. Variability in the use of inflectional morphemes is exceedingly large. At 2;4 the inflectional morpheme score varies between 1 and 39 for those 80 % of children who lie between the 10th and 90th percentile.

Figure 3 (p. 13) shows the developmental curves for the separate inflectional paradigms. Due to the narrow range of the numerical scores for the different inflectional paradigms, where maximums range between 6 and 12, descriptive statistics in terms of means and standard deviation were chosen. The curves in Fig. 3 depict fitted mean values, means, standard deviations and minimum and maximum values are presented in Szagun et al. (2009). The curves show differences in the growth rates of the different inflectional paradigms. Noun plurals and gender marking display the fastest growth. For plurals children reach a mean value of 4.88 out of the maximum score of 6 at 2;6. For gender marking they reach a mean value of 6.16 out of the maximum score of 8. The acquisition of case marking occurs markedly more slowly. It does not exceed a mean score of 1 until 2;0, displays a rapid increase after that and reaches a mean value of 4.31 out of a maximum of 8 at 2;6. All verb forms are acquired more slowly than noun plurals and gender marking. Verb marking on main verbs exceeds a mean score of 1 at 1;11 and then increases to a mean score of 7.11 out of a maximum of 10 at 2;6. The acquisition of modals and forms of the copula is slowest. It exceeds a mean score of 1 at 2;0 and then increases to a mean value of 6.95 out of a total of 12 at 2;6.

Figure 4 (p. 14) shows the growth of sentence complexity. The maximum score is 32. The growth of sentence complexity resembles that of inflectional morphology in that it does not gain momentum until around 2;1 in its central tendency. As indicated by the median curve the inflectional morpheme score increases from 0 at 1;6 to 30 at 2;6 with the largest increases between 2;1 and 2;5. For the 90th percentile the largest increase is earlier, between 1;10 and 2;1, and much later at 2;4 for the 10th percentile. As for inflectional morphology there is a strong floor effect, but a milder ceiling effect. Variability in the growth of sentence complexity is extensive. At 2;4 the sentence complexity score varies between 1 and 30 for those 80 % of children who lie between the 10th and 90th percentile.

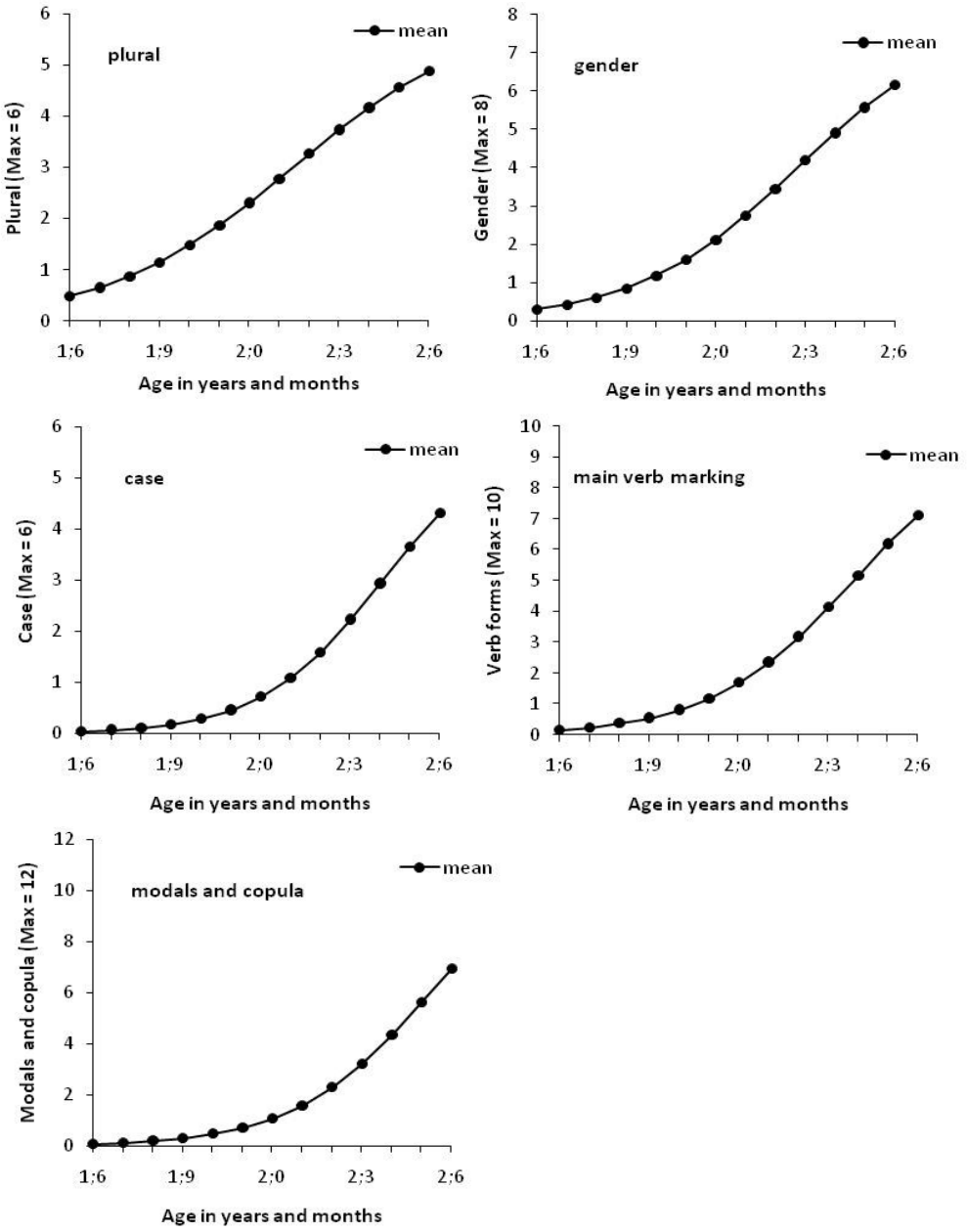

Figure 3. – Fitted mean scores for 5 inflectional paradigms: noun plurals, gender marking, case marking, main verb marking, modals and copula

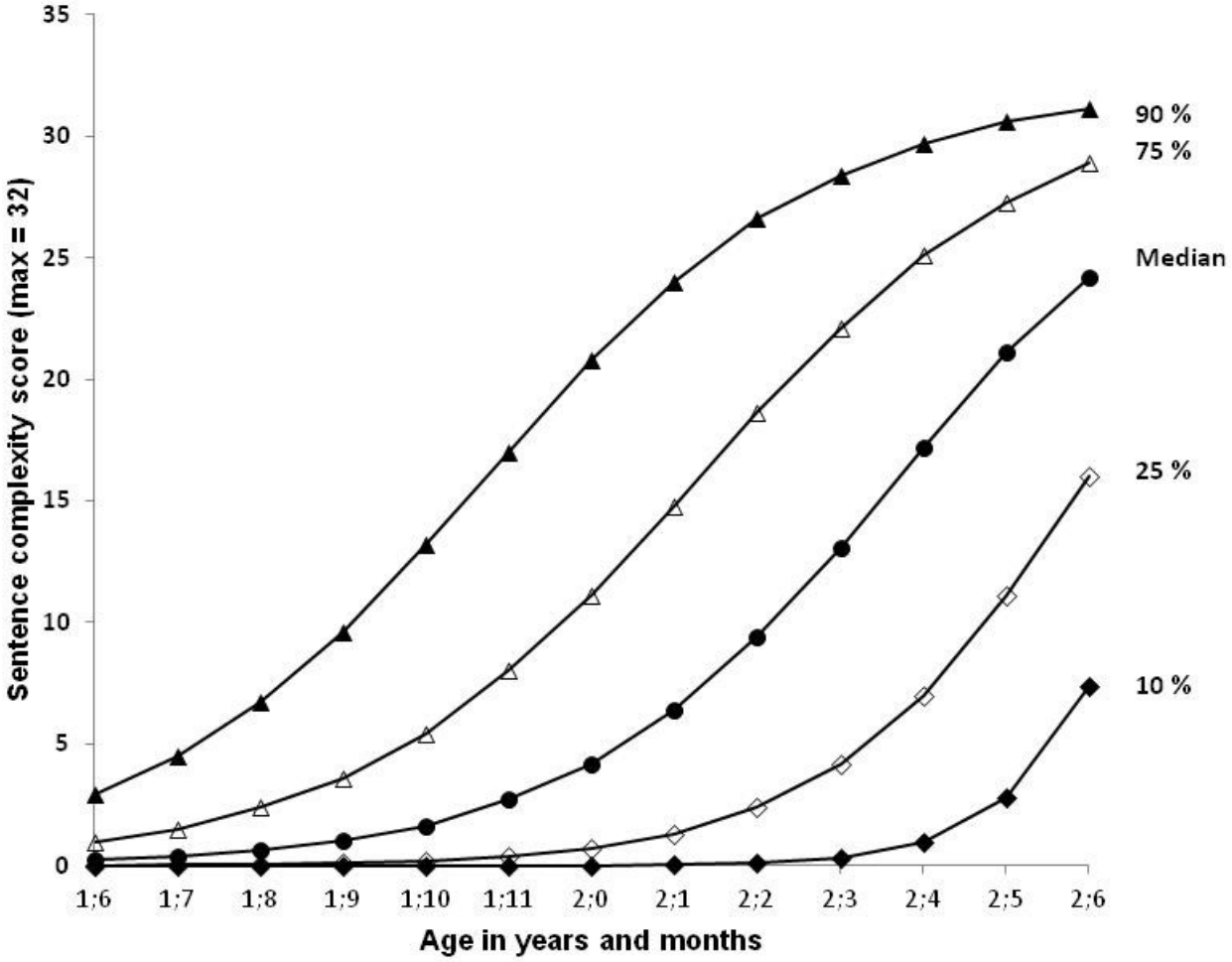

Figure 4. – Fitted percentile scores for sentence complexity

Relations among scales

Relations between vocabulary and grammar scales were investigated by correlational analyses. Bivariate and partial correlations partialing out age were calculated. Table 3 presents the correlation coefficients (Pearson). All correlations are significant ($p < 0.001$) and remain very high, even with age partialed out. The correlations between the two grammar scales are somewhat higher than those between vocabulary and grammar scales. These results indicate that there is a very strong association between vocabulary and grammatical development and that the two aspects of grammar, the use of inflections and of complexity in sentences are even more strongly related.

Table 3: Bivariate and partial correlations between vocabulary, inflectional morphology and sentence complexity

Language measure	Inflectional morphology		Sentence complexity	
	bivariate	partial	bivariate	partial
Vocabulary	0.85*	0.75*	0.85*	0.75*
Inflectional morphology			0.90*	0.84*

* $p < 0.001$

Effects of gender, social class and birth order

The effects of the variables gender, social class and birth order were tested by two-way between-subjects analyses of variance. This was done for each of the major scales: vocabulary, inflectional morphology and sentence complexity.

Gender

Two-way between-subjects analyses of variance with the factors gender (2) and age (13) were calculated for each scale. For vocabulary there was a significant main effect of gender, $F(1,1214) = 44.90$, $p < 0.001$, $\eta_p^2 = 0.036$, and a significant main effect of age, $F(12,1214) = 81.08$, $p < 0.001$, $\eta_p^2 = 0.447$. For inflectional morphology the main effects of gender and age were significant, $F(1,1214) = 34.01$, $p < 0.001$, $\eta_p^2 = 0.027$ and $F(12,1214) = 71.68$, $p < 0.001$, $\eta_p^2 = 0.415$, respectively. For sentence complexity, too, there were significant main effects of gender, $F(1,1214) = 32.94$, $p < 0.001$, $\eta_p^2 = 0.026$, and age, $F(12,1214) = 67.44$, $p < 0.001$, $\eta_p^2 = 0.40$. In no case were the two-way interactions significant. In all cases the main effect of gender favored girls. However, post hoc tests comparing boys and girls per age group revealed significant differences only for some age groups. Girls had significantly higher vocabulary at age 1;9, and a significantly higher inflectional morpheme and sentence complexity score at age 2;5 ($p < 0.05$, Bonferoni adjustment).

Mothers' educational level

Two-way between-subjects analyses of variance with the factors mothers' educational level (4) and age (13) were calculated. For vocabulary there was a significant main effect of educational level, $F(3,1170) = 7.23$, $p < 0.001$, $\eta_p^2 = 0.018$, and a significant main effect of age, $F(12,1170) = 65.60$, $p < 0.001$, $\eta_p^2 = 0.368$. For inflectional morphology there was also a significant main effect of educational level, $F(3,1170) = 8.42$, $p < 0.001$, $\eta_p^2 = 0.021$ and of age, $F(12,1170) = 53.38$, $p < 0.001$, $\eta_p^2 = 0.349$. For sentence complexity there were significant main effects of educational level, $F(3,1170) = 7.34$, $p < 0.001$, $\eta_p^2 = 0.018$, and age, $F(12,1170) = 48.25$, $p < 0.001$, $\eta_p^2 = 0.331$. No two-way interactions were significant. The main effect of educational level was significant for the total sample only. Children of more highly educated mothers had higher levels on all three scales. One difference between the levels of education reached significance: children of mothers with University education had higher inflectional morpheme scores than children of mothers with a middle level of education ($p < 0.05$, Bonferoni adjustment).

Birth order

Children were classified into three groups with respect to birth order: first born (n = 710), second born (n = 378), third + higher birth rank (n = 139). Two-way between-subjects analyses of variance with the factors birth order (3) and age (13) were calculated. For vocabulary there was a significant main effect of birth order, $F(2,1188) = 13.22$, $p < 0.001$, $\eta_p^2 = 0.022$, and a significant main effect of age, $F(12,1188) = 46.59$, $p < 0.001$, $\eta_p^2 = 0.320$. For inflectional morphology there were significant main effects of birth order, $F(2,1188) = 12.10$, $p < 0.001$, $\eta_p^2 = 0.020$, and age, $F(12,1188) = 40.74$, $p < 0.001$, $\eta_p^2 = 0.292$. For sentence complexity there was a significant main effect of birth order, $F(2,1188) = 12.68$, $p < 0.001$, $\eta_p^2 = 0.018$, and of age, $F(12,1188) = 36.93$ $p < 0.001$, $\eta_p^2 = 0.271$. The two-way interactions were not significant. The main effect of birth order was significant for the total sample, and not at any individual age group level. First born children had higher levels on all three language scales. For the grammar scales of inflectional morphology and sentence complexity the difference between first born and second born children was significant ($p < 0.05$, Bonferoni adjustment).

These analyses show that gender, socioeconomic status and birth order have an effect on the growth of early vocabulary and grammar, but the amount of variance they explain is not large. Gender explains between 2.6 % and 3.6 % of the variance in language growth, socioeconomic status between 1.8 % and 2.1 % and birth order between 2.0 % and 2.2 %. By far the largest amount of variance is accounted for by age, between 27.2 % and 44.8 % per analysis.

Psychometric properties

Reliability

Internal consistency

Internal consistency measures the extent to which single items of a scale measure the same content. Several measures of internal consistency were calculated. For vocabulary the score

for each semantic field was correlated with that of each other field and with the total vocabulary score. Of the 231 correlation coefficients (Pearson) between semantic fields, 76 % were above 0.70, 15 % between 0.60 and 0.69, 7 % between 0.50 and 0.59, 1 % between 0.40 and 0.49, and 1 % between 0.30 and 0.39. All correlations below 0.40 and many of those below 0.50 involved the semantic field "sounds effects and animal sounds" for which a less strong relation to words is to be expected. Of the correlation coefficients for the separate semantic fields with the total score 95 % were above 70 % and 5 % between 0.60 and 0.69 and all were significant at the 0.01% level.

For the grammar scales each of the five subscales of the inflectional morpheme scale was correlated with the total inflectional score. Correlation coefficients (Pearson) were between 0.86 and 0.96 ($p < 0.001$).

As further measures of reliability split-half reliability and Cronbach's alpha were calculated for the vocabulary scale. Split-half reliability (using the Spearman-Brown prediction formula) was 0.99, and Cronbach's alpha was 0.99. For the sentence complexity scale Cronbach's alpha was calculated and rendered a coefficient of 0.97. These coefficients were not calculated for the inflectional morpheme scale as this measures different inflectional paradigms and internal consistency on the single item level is not assumed. The results of the correlational analyses are indicative of a high internal consistency of the scales.

Test-retest reliability

Test-retest reliability was tested using the data from the subsample of parents who filled in the questionnaire twice (see *Methods Section*) within a period of seven to 10 days. Of the 60 questionnaires 57 were returned. Correlations between scores on the vocabulary, inflectional morpheme and sentence complexity scales at the first and second data point are presented in Table 4. Even with age partialed out the correlations range between 0.87 and 0.95, and are thus indicative of very good test-retest reliability.

Table 4. - Bivariate and partial correlations between measures of vocabulary and grammar at two data points (n = 57)

Language measure at the 1st data point	Language measure at the second data point	
	bivariate	partial
Vocabulary	0.99*	0.87*
Inflectional morphology	0.97*	0.95*
Sentence complexity	0.96*	0.94*

* $p < 0.001$

Concurrent validity

Concurrent validity was determined by assessing the relation between children's linguistic skills as reported by parents and as measured in the children's spontaneous speech. Data from the subsample of 60 children and parents who took part in the questionnaire study and the recording session of spontaneous speech samples were used for this analysis. One recording had to be discarded for technical reasons. Thus, the sample was reduced to 59. Language

measures derived from spontaneous speech were word type frequency as a measure of vocabulary and MLU as a measure of grammar. Table 5 presents the correlation coefficients (Pearsons) between vocabulary as reported by parents and as measured by word type frequency, between inflectional morpheme use as reported by parents and MLU, and between sentence complexity as reported by parents and MLU. Bivariate correlation coefficients range between 0.90 and 0.93, partial correlation coefficients partialing out age range between 0.82 and 0.85. These values indicate very good agreement between language measures by parent report and language measures based on the child's spontaneous speech.

Table 5. - Bivariate and partial correlations between measures of vocabulary and grammar per questionnaire and on the basis of spontaneous speech (n = 59)

Language measure per questionnaire	Language measure based on spontaneous speech			
	Type frequency		MLU	
	bivariate	partial	bivariate	partial
Vocabulary	0.93*	0.83*		
Inflectional morphology			0.92*	0.85*
Sentence complexity			0.90*	0.82*

* $p < 0.001$

Summarizing the results for tests of reliability and validity, we may conclude that parents report reliably and validly on their child's early vocabulary and of grammatical development.

Short Form Version

In a second step of our research project a short form version of the long questionnaire was developed and empirically tested. The aim was to reduce the vocabulary checklist substantially and reduce the grammar section to just a few questions. Such an instrument can be administered within the time restrictions of many clinical settings.

Item selection and development of the short form

When using a short vocabulary checklist item selection is even more crucial than in a long list, as a child's chance of knowing a word is reduced. Selection of items must therefore be based on empirical evidence of which words discriminate best between slow and fast learners. Items of the short form vocabulary list were drawn from the 600 words of the long form. They were selected along a number of parameters relevant for test construction (Lienert & Raatz, 1998). One was item complexity, as measured by the percentage of children per age group who produce the word. A high percentage indicates that the item is easy, and vice versa. Another was item selectivity which is determined as the correlation of an individual word with the total vocabulary score. It indicates to what extent the differentiation in children with

low and high vocabulary as expressed by an individual item agrees with the total score. Item complexity and selectivity were calculated for 588 words, not counting "sound effects and animal sounds". Items with high selectivity values and wide variability in complexity were selected. As our aim was to create an instrument for all age groups, item choice had to consider an item's value on these parameters per age group. This was a laborious process. We started by selecting appropriate items from the group of children aged 2;0 and then added or deleted items of higher or lower complexity to accommodate all age groups between 1;6 and 2;6 (see Szagun et al., 2009). Early appearing words had to be added to eliminate floor effects at the lower end of the age ranges, and late appearing items had to be added to eliminate ceiling effects at the upper end of the age range. An additional criterion for selection was word class with the aim of having a similar distribution of nouns, verbs, adjectives and "little words", i.e. pronouns, articles, conjunctions, as in the long form. The percentages of nouns, verbs, adjectives and "little words" are shown in Table 6. The short form has a somewhat higher proportion of verbs and a corresponding lower proportion of nouns than the long list. This is due to the fact that the verbs chosen had a higher degree of selectivity and inclusion in the list was decided on the combination of criteria mentioned above. The item selection procedure applied here resembles that of Fenson, Pethick, Renda, Cox, Dale & Reznick (2000).

Table 6. – Percentage of different word classes in the long and short vocabulary list

Type of vocabulary list	% Nouns	%Verbs	% Adjectives	% "little words"
Long form (569 words)*	55	18	9	18
Short form (102)	45	26	12	17

* Words which do not fit into the four categories, i.e. routines, animal sounds, are excluded from the count.

For each child a simulated score on the short vocabulary list was calculated. Data from 1174 children were used, as data collection for the norming study with all 1240 participants had not yet been completed. Correlations between simulated scores and original vocabulary scores were calculated repeatedly until no further improvement through word selection could be obtained and the final 102 words were selected. Children's simulated short-form scores on the list containing 102 words and children's original vocabulary scores were correlated per age group and for the total sample. The correlations (Pearson) per age group ranged between 0.989 and 0.998, the correlation for the sample as a whole with age partialed out was 0.992 ($p < .001$).

Description of the instrument

The short form version is designed for children between 1;6 and 2;6. It contains a vocabulary list of 102 words. Five words are added and integrated into the list bringing the total presented to 107 words. These are the German equivalents of *mummy, daddy, granny, ball* and *no* which were amongst the easiest words for all children. They serve to give parents the opportunity for a "yes" answer in case their child does not produce any of the 102 words

of the list which does not contain any sound effect 'words'. They are not counted in the total score. In the interest of brevity the short form contains only three questions relating to grammar, a question asking if the child already produces word combinations and two questions on inflectional morphology asking about the use of noun plurals and of gender marking. The three questions are presented in same way as in the long questionnaire using a "not yet", "sometimes" and "often" format. The question about word combinations is fundamental to developing grammars (Szagun et al., 2009). The two questions relating to inflectional morphology were selected because they correlated best with the long inflectional morpheme scale and the respective subscales (see Szagun et al., 2009). The short version contains the same basic information sheet as the long version of the questionnaire.

Description of empirical study and participants

An empirical study was carried out with a number of aims. The first one was to check if the simulated vocabulary scores concord with vocabulary scores obtained when parents fill out the short vocabulary list. Parents may respond differently to a short list of words (Fenson et al., 2000). It is therefore necessary to check empirically if the total vocabulary score differs if the 102 words are presented on their own or within a list of 600 words. Secondly, we wanted to check empirically if the long and the short vocabulary list measure the same phenomenon, i.e. children's vocabulary size. Therefore, scores on both lists need to be obtained from the same participants and examined with respect to their extent of agreement. Thirdly, to control for an order of items effect a second version of the short word list with items in a different order was presented to parents and results on the two versions compared.

Participants in the empirical investigation were 411 parents of children aged between 1;6 and 2;6. The sample was drawn from the same paediatrician's practices in the same towns as the norming sample for the long questionnaire. The sample was very similar to this large norming sample of 1240, and therefore sample characteristics are not described in further detail here (see *Methods Section: Participants* and Szagun et al., 2009). A sample of 356 parents filled in the short form of the questionnaire. Of these a subsample of 113 were sent the long form and the short form of the questionnaire a week apart. Sixty parents filled in the short form first and the long form thereafter, for 53 parents this was done in the reverse order. Another 55 parents were sent a version of the short questionnaire with items in a different order.

Results

In order to economise on sample size a sequential design ("Triangular Sequential Design", Rasch & Kubinger, 2004) was used which assessed when sample size was sufficient to decide that there was no difference between the empirical vocabulary scores and the simulated scores. In applying the method of triangular sequential testing, expected means per age group and extent of variability allowed have to be specified. These values were specified on the basis of the results with the long questionnaire, as means and standard deviations per age group are known. We allowed a deviation from the mean of 2/3 of the standard deviation. The procedure controls for the occurrence of Type I and Type II errors and sets α at 0.05 and β at 0.20. Questionnaire results were entered into the calculations sequentially in the order in which the questionnaires were returned. Empirical data collection was ended when there were sufficient questionnaires in order to make a decision per age group whether scores differed

from the simulated scores or not. For all 13 age groups the sequential triangular test allowed a decision supporting the hypothesis that the simulated mean vocabulary scores did not differ from the empirically derived mean vocabulary scores. For this decision between 12 and 35 questionnaires per age group were necessary in the empirical study. Altogether 356 questionnaires were used. In addition to the sequential triangular test for each group the simulated scores were compared to the empirical scores by t-test for independent samples. All t-tests were non-significant. On the basis of these tests we conclude that the simulated scores and empirically derived scores do not differ. We are therefore justified to use the simulated scores based on a sample of 1174 participants as norm data. Figure 5 shows the growth of vocabulary between 1;6 and 2;6 as measured by the short vocabulary list. In its central tendency, as indicated by the median, vocabulary increases from 5 words at 1;6 to 85 words at 2;6. At 2;0 the number of words of those 80 % children who lie above the 10th and below 90th percentile ranges between 5 and 80. The variability in vocabulary growth measured with the short vocabulary list parallels that measured with the long vocabulary list and is very substantial. Results with the short list show a floor effect and a more pronounced ceiling effect than for results with the long list.

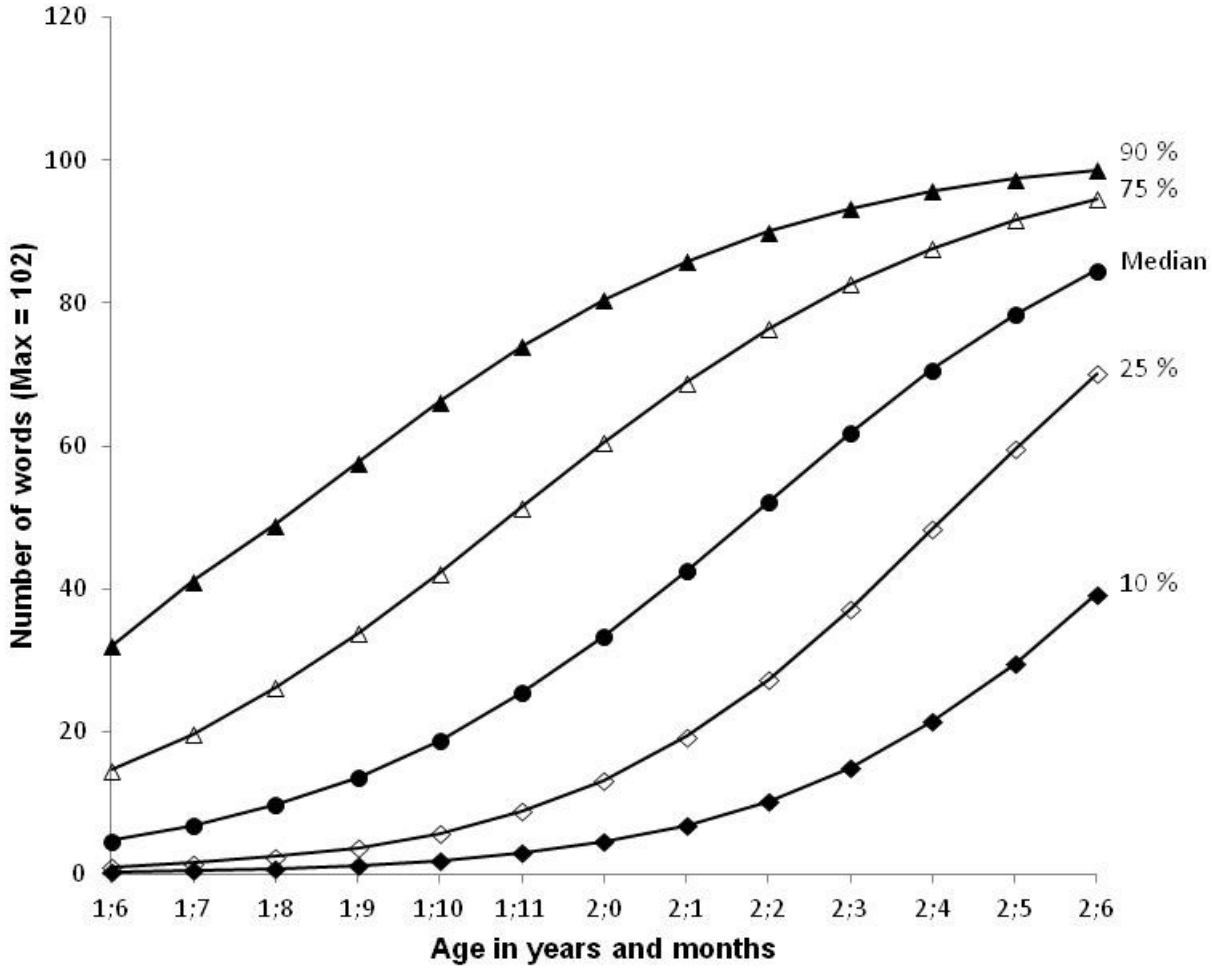

Figure 5. – Fitted percentile scores for words of the Short Form Version

A possible order effect was tested in the following way. The scores of the 55 questionnaires with items in a different order were compared with scores from 55 questionnaires with items in the standard order. These 55 questionnaires were randomly drawn from the sample of 356 except that age and gender corresponded. There was no significant difference between the vocabulary scores of the two groups (independent t-test). We conclude that the order of presentation of words in the short list does not affect parents' response behaviour.

As for the long form version, we are focusing here on the main scales of the questionnaire and are not presenting the results for the three questions on grammar. The results for the effects of gender, socioeconomic status and birth order in the sample of 456 participants of the new empirical study were similar to those for the norming sample and will not be presented here (for details on these issues, see Szagun et al., 2009).

Psychometric properties

Reliability

As measures of reliability split-half reliability and Cronbach's alpha were calculated. This was done with the simulated scores and the sample of 1174 participants. Split-half reliability (using the Spearman-Brown prediction formula) was 0.95, and Cronbach's alpha was 0.99. The short vocabulary list of 102 items is characterized by a high degree of internal consistency.

Concurrent validity

The lack of a significant difference between the simulated vocabulary scores and the empirically derived vocabulary scores can be taken as evidence for the concurrent validity of the simulated scores which serve as our norm. Another test of concurrent validity of the short vocabulary list was carried out by correlating the scores obtained on the short list with the scores obtained on the long list in the sample of 113 parents. Correlations (Pearson) between the two sets of scores were 0.96, and 0.94 when age was partialed out ($p < 0.001$). These results show that the short vocabulary list is a valid instrument for assessing young children's vocabulary.

Discussion

The present results provide the first normative data on the typical course and extent of variability in lexical and grammatical development in German-speaking children between 1;6 and 2;6. The observed variability in the growth of vocabulary and grammar is so large that children at the same language level can differ by up to 12 months of age. This encompasses those 80 % of the children who lie between the 10^{th} and 90^{th} percentile of the distribution, not counting children at the extreme ends.

The growth of vocabulary accelerates before grammar. Focusing on the central tendency vocabulary growth begins to accelerate around 1;9, whereas inflectional morphology and sentence complexity do not show marked increases in growth until around 2;1. Results for the subscales of inflectional morphology show that the acquisition of noun plural marking and

gender marking progresses most rapidly. This is followed by the acquisition of case marking and main verb marking, while the acquisition of modal verbs and copula occurs most slowly.

The different linguistic skills are strongly related, even when age is controlled for. Gender, birth order and socioeconomic status have a significant effect on the acquisition of vocabulary and grammar, though the size of the effects is small. The effects of the demographic factors indicate that girls, children of more highly educated mothers and first born children display faster linguistic growth.

Our German CDI adaptation demonstrates very good reliability in terms of internal consistency and test-retest reliability. It also demonstrates very good concurrent validity as documented by the very high correlations between questionnaire scores and measures of vocabulary and grammar obtained on the basis of spontaneous speech.

A short form version of our German CDI adaptation was developed on the basis of the empirical results with the long form. It contains a vocabulary list of 102 words and three questions on grammar. Criteria for selection of words were broad variability of item complexity, high item selectivity, and word class. Simulated vocabulary scores based on the large sample are used as norm data, as they do not differ from empirical scores with a new, but smaller sample. The short vocabulary list demonstrates very good reliability as measured by internal consistency. Concurrent validity, as expressed by the correlations between the long form and the short form is very high. The extent of variability in vocabulary growth as measured by the short vocabulary list parallels that measured with the long list.

The extensive variability observed in our data converges with the results for American English and other languages using adaptations of the CDI, such as Italian, Hebrew, Swedish, British English, Dutch and Danish ((Fenson et al., 1994, Fenson et al., 2007; Caselli et al., 1999; Maital et al., 2000; Berglund & Eriksson, 2000; Hamilton et al., 2000; Zink & Lejaegere, 2002; Bleses et al., 2008 a, b). The present norm data for the early development of German add to the evidence that extensive individual variability, not a fixed maturational timetable, is characteristic of early language development. Extensive individual variation rather than fixed ages for certain linguistic achievements may, indeed, turn out to be a universal in early language development.

A special characteristic of our German CDI adaptation is its very detailed section on inflectional morpheme acquisition. In keeping with the aim of the CDIs to assess language-specific behaviours in the toddler period, this scale addresses the inflectional paradigms of noun plural, gender and case marking, verb inflection and auxiliaries which, typically, begin to develop during this period. The results are highly informative. They allow a fairly detailed assessment of the level of inflectional morpheme acquisition in the different paradigms at a particular age. They also show a developmental sequence in which noun plurals and gender marking are acquired first, followed by case marking and main verb markings and finally by modals and the copula. This sequence converges with evidence based on spontaneous speech data from longitudinal studies (Clahsen, 1982, 1984; Szagun, 2001; Behrens, 1993, 2002; Szagun et al., 2007). Noun plurals and gender marking on articles are amongst the earliest morphemes acquired by young German-speaking children (Szagun, 2001; Behrens, 2002; Szagun et al., 2007). Case marking and verb marking for present tense and on the past participle develop somewhat more slowly (Clahsen, 1982, 1984; Szagun, 2004) with a close co-occurrence of case marking and verb person marking Clahsen (1982, 1984). Forms of modals, auxiliaries and the copula are acquired notably more slowly (Behrens, 1993). The

present results provide the first evidence for such a sequence in a normative sample. However, as the data are cross-sectional they only tells us that for the population as a whole this particular sequence is observed, but not if it is observed in each individual of the population. To examine if this is the case further analyses are required for which the present set of norm data can be used. Scaling analyses can be performed to find out if and to what extent one grammatical paradigm is mastered before another – equivalent to the analyses for lexical items performed with the CLEX database (Jørgensen, Dale, Bleses & Fenson, 2010).

Regarding the influence of demographic factors on early language development in our data the effect of gender is strongest, whereas socioeconomic status and birth order exert a somewhat milder influence. This resembles the pattern observed for the USA, Northern Belgium and Denmark (Fenson et al., 1994; Fenson et al., 2007; Zink & Lejaegere, 2002; Bleses et al., 2008b). However, the amount of variance explained by demographic variables is slightly larger in the German normative sample. This may well be due to child-rearing environments in Germany where most children under three years are brought up in family environments with the mother as the main caregiver. Such an environment is likely to intensify the influence of such variables as socioeconomic status, gender and birth order. As the influence of socioeconomic class is mediated by maternal language input (Hoff, 2003), having the mother as the main source of language input would heighten the effects of the quality of her speech. Children of more highly educated mothers who produce rich language input would benefit and vice versa. In first born children the positive effect of having an adult as the main dialogue partner would also be heightened in a family environment with one adult as the main caretaker. Finally, in as far as gender-role typical behaviour contributes to girls' advantage in language development (Bornstein et al., 2004), a social environment with the mother as the main interaction partner would enhance this effect. This view is supported by the lack of a gender effect in the Swedish and Israeli CDI studies (Berglund & Eriksson, 2000; Maital et al., 2000). In both societies nursery environments for children younger than three years are more widely spread than in Germany.

A limitation of the present study is that the highest educational levels are over-represented in our sample when compared to the census data for the population of women between 20 and 40 years. This is the case in many studies with the CDI (Fenson et al., 1994; Hamilton et al., 2000; Bleses et al., 2008b), and, in fact, our sample compares quite favourably with other CDI samples. The method of recruiting participants which relies on anonymous and voluntary participation may lead to an overrepresentation of more highly educated parents. However, our sample was diverse enough to allow the examination of the effect of socioeconomic status. As this effect was not significant per individual age group, we regard the norms per monthly age group as adequate but would urge caution when classifying border line cases from lower socioeconomic backgrounds as language delayed (see Szagun et al. 2009).

In creating the short form version of the present questionnaire great care was taken to ensure that those items were selected which discriminate best between children with fast and slow language development. During the selection process items were constantly modified and the simulated vocabulary scores based on the short lists tested for their fit with the scores based on the long list. The results of this process and the fit between items per age group are presented in detail in Szagun et al. (2009). The procedures of item selection and checking the simulated short form scores for agreement with long form scores resembled that of Fenson et al., (2000). Where our procedure differs is that we use the simulated scores as norm data. This

decision is based on the convergence of simulated and empirically derived scores when stringent criteria were used in testing for differences between the two sets of scores. When using the triangular sequential method (Rasch & Kubinger, 2004) we allowed only 2/3 of a standard deviation from the mean. We further ascertained that there was no difference by a series of t-tests for each age group. Taken together, this provides adequate justification for the decision to regard the simulated scores as norm data.

Finally, we would like to discuss the use of our CDI adaptation in basic research and for clinical purposes. Both questionnaires, the long and the short form version, provide an excellent means for matching children on language level. Which version is used would depend on how much information on language status the researcher wants. Detailed information on grammatical status, for instance, can only be obtained from the long version. In our view, for much research matching on language level is more appropriate than the predominant method of matching on age which ignores the large age differences between children at the same language level. For cross-linguistic research our data allow a variety of investigations concerning comparisons of speed of lexical and grammatical growth. Our lexical norm data have been added to the CLEX data base (Jørgensen et al., 2010). This enables researchers to obtain norms for the acquisition of individual words which can be used in cross-linguistic comparisons or as prerequisite information when designing language experiments with young children.

The questionnaires can also be used for research with atypical populations. We established the validity and reliability of our CDI adaption for assessing spoken language development in deaf children with cochlear implants and have used the long form to track the lexical and grammatical development of a large sample of cochlear-implanted children (Szagun, 2010). While the age norms cannot be applied to such children, their language development can be monitored by repeated administration of the questionnaire several months apart. For American English, too, the CDI has been shown to be a valid and reliable measure of cochlear-implanted children's spoken language abilities (Thal, DesJardin & Eisenberg, 2007). The questionnaire results allow a first impression of areas of particular difficulties or hint at possible differences in atypical populations (Bates & Goodman, 1999). In the case of German-speaking children with cochlear implants the questionnaire results on inflectional morpheme use gives a first and rough impression of the specific difficulties these children have with case and gender marking on articles (Szagun, 2004; 2010).

Our norming data provide the basis for developing criteria for what may constitute language delay in German-speaking toddlers which are based on the variability typically observed in the population. In accordance with studies using the CDI (Bates, Dale & Thal, 1995; Fenson et al., 2007) we recommend to classify children who score in the lowest 10 % of their age group as children with late onset of language. Such late onset should be viewed neutrally and not per se seen as a risk for subsequent language impairment. There are a variety of reasons why a child may be slow in developing language, including the child's temperament and environmental factors (Bates, Bretherton & Snyder, 1988; Bates et al., 1995; Hoff, 2003), and, at present, there is insufficient evidence to predict subsequent language impairment from the late onset of language in one- to two-year olds (Nelson, Nygren, Walker & Panoschka, 2006; IQWiG, 2007). In the German situation where language assessment is part of preventive paediatric health checks between 1;6 and 3;0 a monitoring strategy may be appropriate (Paul, 1996; Szagun et al., 2009). Our questionnaires are well suited to this

purpose. The short form version can be use within the time limitations of a paediatric consultation. We recommend the use of the long form if the child continues to score in the lowest decentile at 2;6, as this version renders fairly detailed information on grammatical development and may thus point to specific difficulties in this area (for more details see Szagun et al., 2009).

The present norm data for the early development of German add to the evidence that extensive individual variability, not a fixed maturational time-table, is characteristic – and possibly universal - in early language development. Our CDI adaptation can be used in basic research and for clinical assessment alike. In the current situation of assumed wide-spread language delay in German-speaking children, an instrument which allows the placement of an individual child relative to population norms, would seem a valuable tool for language assessment.

References

Abbot-Smith, K. & Behrens, H. (2006). How know constructions influence the acquisition of new contructions: the German periphrastic passive and future constructions. *Cognitive Science, 30,* 995-1026.

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

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*, pp. 96-152. Oxford: Basil Blackwell.

Bates, E. & Goodman, J. (1999). On the emergence of language from the lexicon. In B. MacWhinney (Ed.), *The emergence of language,* pp. 29-79. Mahwah, N.J.: Erlbaum.

Behrens, H. (1993). *Temporal reference in German child language*. Amsterdam: Wöhrmann.

Behrens, H. (2002). Learning multiple regularities: Evidence from overgeneralization errors in the German plural. A. H.-J. Do, L. Dominguez & A. Johanse (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development,* pp. 72-83. Somerville/MA: Cascadilla Press.

Berglund, E. & Eriksson, M. (2000). Communicative development in Swedish children 16-28 months old: The Swedish early communicative development inventory – words and sentences. *Scandinavian Journal of Psychology, 41*, 133-144.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomson, P., Madsen, T. & Basbøll, H. (2008a). Early vocabulary development in Danish and other languages. *Journal of Child Language, 35*, 619-650.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomson, P., Madsen, T. & Basbøll, H. (2008b). The Danish Communicative Developmental Inventories: validity and main developmental trends. *Journal of Child Language, 35*, 6519-669.

Bockmann, A. & Kiese-Himmel, D. (2006). *ELAN - Eltern Antworten. Elternfragebogen zur Wortschatzentwicklung im frühen Kindesalter. Manual.* (*ELAN – Parent answers. Parental questionnaire for vocabulary development in early childhood. Manual*) Göttingen: Test Zentrale.

Bornstein, M., Hahn, C.-S. & Haynes, M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24*, 267-304.

Brandt, S. Diessel, H. & Tomasello, M. (2008). The acquisition of German relative clauses: A case study. *Journal of Child Language, 35*, 325-348.

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

Clahsen, H. (1982). *Spracherwerb in der Kindheit: Eine Untersuchung zur Entwicklung der Syntax bei Kleinkindern.* (*Language acquisition in childhood: An investigation into the development of syntax in young children*). Tübingen: Narr.

Clahsen, H. (1984). Der Erwerb von Kasusmarkierungen in der deutschen Kindersprache (The acquisition of case marking in German child language). *Linguistische Berichte, 89,* 1-31.

Clark, E. (2003). *First Language Acquisition*. Cambridge University Press.

Fenson, L., Dale, P., Reznick, J. S., Bates, E., Thal, D. & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*.

Fenson, L., Marchman, V., Thal, D., Dale, Ph., Reznick, J.S. & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual*, 2nd edition. Baltimore: Brookes Publishing.

Fenson, L., Pethick, S., Renda, C., Cox, J.L., Dale, P.S. & Reznick, J.S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21* (1), 95-116.

Grimm, H. & Doil, H. (2006). ELFRA – *Elternfragebögen für die Früherkennung von Risikokindern (Parental questionnaire for the early identification of children at risk).* Göttingen: Hogrefe, 1. Auflage/1st edition, 2000.

Hamilton, A., Plunkett, K. & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language, 27,* 689-705.

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*, 1368-1378.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Economical Management in the Health Care System) (2007). Früherkennungs-untersuchung auf umschriebene Entwicklungstörungen des Sprechens und der Sprache (Investigation into the early identification of specific language impairment). Berichtsplan (Report) S06-01. Köln: IQWiG.

Jørgensen, R., Dale, P., Bleses, D & Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. *Journal of Child Language, 37,* 419-428.

Kauschke, C. & Hofmeister, E. (2002). Early lexical development in German: a study on vocabulary growth and vocabulary composition during the second and third year of life. *Journal of Child Language, 29*, 735-757.

Kern, S. (2007). Lexicon development in French-speaking infants. *First Language, 2*, 191-212.

Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (*Test construction and test analysis*). Weinheim: Beltz.

López Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S. & Martínez, M. (2005). Inventarios de Desarollo Comunícativo MacArthur: Maual Técnico, Madrid: Ediciones TEA.

Maccoby, E. (1966). Sex differences in intellectual functioning. In E. Maccoby (Ed.), The development of sex differences (pp. 25-55). Stanford, CA: Stanford University Press.

MacWhinney, B. (2000). MacWhinney, B. (2000). *The CHILDES Project: tools for analyzing talk,* 3rd edition. Hillsdale, New Jersey: Erlbaum.

MacWhinney, B. (2009). Talk Bank. http://childes.psy.cmu.edu

Maital, S., Dromi, E., Sagi, A. & Bornstein, M. (2000). The Hebrew communicative development inventory: language specific properties and cross-linguistic generalizations. *Journal of Child Language, 27*, 43-67.

Mills, A. (1985). The acquisition of German. In D.Slobin (Ed.), The *cross-linguistic study of language acquisition*, pp 141-254. Hillsdale N.J.: Erlbaum.

Nelson, H.D, Nygren, P, Walker, M. & Panoschka, R. (2006). Screening for speech and language delay in preschool children: systematic evidence review for the US Preventive Services Task Force. *Pediatrics, 117 (2)*, e298-e319.

Paul, R. (1997). Understanding language delay: A reply to van Kleeck, Gillam, and Davis. *American Journal of Speech-Language Pathology, 6 (2),* 40-49.

Rasch, D. & Kubinger, K.D. (2004). *TRIQ Handbuch – Versuchsplanung und Auswertung von Sequentiellen Dreieckstest (Triangular Sequential Designs) Version 1.31 (TRIQ Handbook – design and analysis of triangular sequential tests)*. Rostock: BioMath GmbH.

Rescorla, L. (1989). The Language Development Survey: A Screening tool for delayed language in toddlers. *Journal of Speech and Language disorders, 54 (4),* 587-599.

Statistisches Bundesamt (2006). *Statistisches Jahrbuch für die Bundesrepublik Deutschland 2006 (Statistical Yearbook for the Federal Republic of Germany 2006)*. Wiesbaden.

Stromswold, K. (2000). The cognitive neuroscience of language acquisition. In M.S. Gazzaniga (Ed.), *The new cognitive neurosciences*, pp. 909-932.Cambridge, MA: MIT Press.

Szagun, G. (2001). Learning different regularities: The acquisition of noun plurals by German-speaking children. *First Language*, *21*, 109-141.

Szagun, G. (2004a). Learning by ear: On the acquisition of case and gender marking by German-speaking children with cochlear implants and with normal hearing. *Journal of Child Language, 31*, 1-30.

Szagun, G. (2004b). German – Szagun. Talk Bank. http://childes.psy.cmu.edu

Szagun, G. (2010). Einflüsse auf den Spracherwerb bei Kindern mit Cochlea Implantat: Implantationsalter, soziale Faktoren und die Sprache der Eltern (Influences on the acquisition of language in children with cochlear implants: age of implantation, social factors and parental language). *hörgeschädigte kinder – erwachsene hörgeschädigte (hearing- impaired children – adult hearing-impaired persons), 47 (1)*, 8-36.

Szagun, G., Steinbrink, C., Franik, M. & Stumper, B. (2006). Development of vocabulary and grammar in young German-speaking children assessed with a German language development inventory. *First Language, 26, 259-280*.

Szagun, G., Stumper, B. & Schramm, A.S. (2009). *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform) (Questionnaire on early language development (FRAKIS) and short form (FRAKIS-K).* Frankfurt: Pearson Assessment.

Szagun, G., Stumper, B. Sondag, N. & Franik, M. (2007). The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language, 34,* 445-471.

Thal, D., DesJardin, J. & Eisenberg, L. (2007). Validity of the MacArthur-Bates Communicative Development Inventories for measuring language abilities in children with cochlear implants. *American Journal of Speech-Lanugage Pathology, 16,* 54-64.

Von Suchodoletz, W. (2008). Der SBF-2-KT – ein neuer Kurztest zur Früherkennung von Late Talkers bei der U7 (The SBF-2-KT – a new short test for the identification of late talkers at the U7 examination). *Kinderärztliche Praxis*, 308.

Zink, I. & Lejaegere, M. (2002). *N-CDIs: Lijsten voor Communicative Ontwikkeling*. Leuven/Leusden: Acco.

Acknowledgements

This research was funded by Deutsche Forschungsgemeinschaft (German Science Foundation), grants No. Sz 41/11-1 and No. Sz 41/11-2 to the first author. We are most grateful to the children and their parents who so willingly participated in this study. Mohsen Haj Bagheri, Melanie Franik, Nina Sondag, Tim Oesterlau and Sarah Deutscher helped with data collection and management. Special thanks go to the paediatrians, the assistant staff in their practices, and the staff in the day care centres whose co-operation made this study possible.