The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities*

GISELA SZAGUN, BARBARA STUMPER,
NINA SONDAG AND MELANIE FRANIK

Department of Psychology, University of Oldenburg, Germany

ABSTRACT

The acquisition of noun gender on articles was studied in a sample of 21 young German-speaking children. Longitudinal spontaneous speech data were used. Data analysis is based on 22 two-hourly speech samples per child from 6 children between 1;4 and 3;8 and on 5 two-hourly speech samples per child from 15 children between 1;4 and 2;10. The use of gender marked articles occurred from 1;5. Error frequencies dropped below 10% by 3;0. Definite and indefinite articles were used with similar frequencies and error rates did not differ in the two paradigms. Children's errors were systematic. For monosyllabic nouns and for polysyllabic nouns ending in -*el*, -*en* and -*er* errors were more frequent for nouns which did not conform to the rule that such nouns tend to be masculine. Furthermore, children erred in the direction of the rule overgeneralizing *der*. Correct gender marking was also associated with adult frequency of noun use. The present data is evidence for the early use of phonological regularities of noun structure in the acquisition of gender marking.

INTRODUCTION

Noun gender assignment in German is a puzzling phenomenon, and a tough one for non-native speakers. Not so long ago the view was widespread that there were too few discernable regularities to grant gender classification in German the status of an orderly system, and it was viewed as arbitrary (Maratsos, 1982). After all, why should *der tisch* 'the table' be masculine,

TABLE 1. *German definite and indefinite articles[a] in the singular*

Case	Masculine		Feminine		Neuter	
	Definite	Indefinite	Definite	Indefinite	Definite	Indefinite
Nominative	der	ein	die	eine	das	ein
Accusative	den	einen	die	eine	das	ein
Dative	dem	einem	der	einer	dem	einem
Genitive[b]	des	eines	der	einer	des	eines

[a] For indefinite articles there may be different phonetic realizations of a particular article form, as contracted forms are frequent in spoken language (for details see Szagun, 2004a).
[b] The genitive has not been reported in child language.

die strasse 'the street' be feminine and *das pferd* 'the horse' be neuter? Grammatical gender may even override natural gender, as in the case of the neuter noun *das mädchen* 'the girl'. However, some researchers deviated from the majority view of arbitrariness in German gender assignment and discovered regularities that hold for at least part of the vocabulary (Altmann & Raettig, 1973; Köpcke & Zubin, 1983, 1984). Surprisingly, too, children acquiring German as their native language do not appear to be too troubled by gender assignment, at least as far as we know from a limited set of data (Mills, 1985, 1986). How do young children manage to acquire a system – or non-system, as the case may be – which is so difficult for adult language learners?

There are three gender paradigms in German: masculine, feminine and neuter. Gender marking is pervasive and affects many parts of speech. The most frequent ones are definite and indefinite articles. Besides articles, gender is marked on all other determiners, relative and question pronouns, personal pronouns in the third person and on attributive adjectives. The present analysis will focus on gender marking on articles, which is most frequent. Table 1 presents the forms of the definite and indefinite articles. Case marked forms are included, as gender marking is intertwined with case marking in such a way that some forms confound the gender paradigms. Thus, in the definite paradigm the form *der* denotes nominative masculine and dative feminine, and the form *dem* is dative masculine and neuter. Plural forms are not presented, as plurals are not marked for gender.

Three types of regularities of gender assignment have been distinguished for German: semantic, morphological and phonological (Altmann & Raettig, 1973; Köpcke, 1982; Köpcke & Zubin, 1983, 1984; Heidolph, Flämig & Motsch, 1984; Mills, 1986; MacWhinney, Leinbach, Taraban & McDonald, 1989; Duden, 1995). Most of these regularities are not deterministic rules, but are of a probabilistic nature. In other words, nouns containing a certain phonological or morphological pattern, or belonging to a particular semantic

446

field, display the tendency to be associated with a particular gender (Altmann & Raettig, 1973; Köpcke & Zubin, 1983, 1984). Köpcke & Zubin (1984) present a very complex analysis of semantically based gender assignment. Some of their categories are also found in standard grammars (Heidolph *et al.*, 1984; Duden, 1995). Only some categories will be given here. There is a tendency for nouns denoting minerals, weather, birds and beverages to be masculine, a tendency for nouns denoting trees, flowers, knowledge and crafts, and musical instruments to be feminine, and a tendency for chemical elements, metals, superordinate animal names and other superordinates to be neuter (Köpcke & Zubin, 1984; Heidolph *et al.*, 1984, see also Mills, 1986). In many cases there is also a correspondence between natural gender and grammatical gender (Mills, 1986). However, the semantic principle interacts with morphological and phonological regularities and may apply only when these latter do not (Köpcke & Zubin, 1984; Mills, 1986).

Morphological regularities and regularities of word endings are treated in conjunction in a number of classification attempts (Heidolph *et al.*, 1984; MacWhinney *et al.*, 1989; Duden, 1995). Heidolph *et al.* (1984) use the term 'word formation rules' for word endings in monomorphemic words and for derivational suffixes which are associated with a particular gender. Word endings associated with masculine gender are *-en*, *-er* and *-el* (Heidolph *et al.*, 1984; MacWhinney *et al.*, 1989). A number of derivational suffixes are associated with feminine gender, such as *-heit*, *-keit*, *-ung*, *-schaft* and *-ei* forming abstract nouns, others with neuter gender, such as the diminutive suffixes *-chen* and *-lein* and the prefix *ge-*added to another noun meaning 'collective' (Heidolph *et al.*, 1984; Mills, 1986; MacWhinney *et al.*, 1989; Duden, 1995). Another morphological rule is the 'Last-Member Priniciple' (Köpcke & Zubin, 1984), which states that compound nouns are assigned the gender of the final segment.

Köpcke (1982) developed a classification system based on phonological regularities, which holds for monosyllabic nouns and is not found in standard grammars (Heidolph *et al.*, 1984; Duden, 1995). In a detailed vocabulary analysis of adult language Köpcke (1982; Köpcke & Zubin, 1983, 1984) was able to show associations of word endings and word initial sounds with gender, mostly masculine, for monosyllabic nouns. Thus, monosyllabic words beginning with [kn], [dr] or [tr], with [ʃ]+consonant, and those ending on a nasal+consonant, as well as those with two initial and two final consonants, tend to be masculine. Köpcke & Zubin (1984) also subsume these five regularities under one regularity which states that the more consonants in word initial and/or word final position the higher the probability of masculine gender assignment. Some monosyllabic nouns, however, are associated with feminine gender. These are a small number of words ending in [ft], [xt], [çt] and in [u:r] or [y:r] (Köpcke & Zubin, 1983). Mills (1986)

447

refers to two more phonological patterns in polysyllabic words which are strongly associated with feminine gender. These are words ending in -*e* (schwa) and words ending in -*ie*. Köpcke & Zubin (1983) present empirical evidence that adults make use of the specified phonological regularities when assigning gender to monosyllabic nonsense words. Thus, phonological patterns have psychological reality and are one principle guiding gender assignment in German (Köpcke & Zubin, 1983, 1984).

What is known about children learning gender assignment? As mentioned before, forms which carry gender marking are determiners and adjectives. The most frequent forms are articles. Available data on children's acquisition of gender marking are mainly based on nineteenth- and early twentieth-century diary studies of individual children (summarized in Mills, 1986) and observational data from three children (Mills, 1986). Additionally, analyses of the acquisition of case marking (Clahsen, 1984; Szagun, 2004a) provide some information on the acquisition of gender. While Mills (1986) reports the sporadic use of gender marked articles shortly after 2;0 and frequent use from 2;4, Clahsen (1984) does not observe article use until ages 2;9 and 2;10. In the data analyzed by Mills (1986), indefinite articles were used earlier and more frequently than definite articles. Error rates were higher for indefinite forms but dropped to a negligible level by age 3;0. In the definite paradigm errors were altogether rare. The most frequent error was an overgeneralization of *die*. In Szagun's (2004a) analysis, which is based on an extensive data set of the spontaneous speech of six children between 1;4 and 3;8, the more frequent use of indefinite articles was confirmed for the nominative case at MLU levels 1·9, 2·8 and 3·8 but not thereafter. However, errors of gender in the nominative were not more frequent for indefinite than for definite articles. Concerning the definite article, *die* was a frequent error, but not significantly more frequent than the erroneous use of *der*.

Only Mills (1986) raises the question of how children's correct and incorrect use of gender marking relates to the phonological forms of nouns. For observational data she observed that the children's vocabulary contained many feminine words ending in -*e* and that these were used with correct gender. Thus, high frequency and consistency in gender marking for words ending in -*e* in the language at large may contribute to the early acquisition of correct gender marking for these words. Mills (1986) also notes erroneous gender marking like *die hund* 'dog' and *die bagger* 'dredger' for words associated with masculine gender according to phonological and word ending regularities (Köpcke & Zubin, 1984; Heidolph *et al.*, 1984), concluding that the particular phonological rules are not yet present. In an experiment with seven- to nine-year-olds, Mills (1986) tested whether children make use of the phonological regularities specified by Köpcke & Zubin (1984) when assigning gender to monosyllabic nonsense words. This

448

was confirmed for masculine but not feminine words. Thus, there is evidence that older children make use of phonological regularities when they assign noun gender. But, unlike adults, they do so for masculine monosyllabic nouns only (Mills, 1986).

These existing analyses on the acquisition of gender marking diverge on some issues and leave others unanswered. There are diverging results on age of onset of gender marking and on erroneous gender marking of indefinite and definite articles. These divergences are likely to be due to differences in sampling procedures (see Szagun, 2004a) or levels of analysis. Mills' (1986) and Clahsen's (1984) analyses are based on small speech samples of individual children. The delayed age of onset of gender marking in Clahsen's (1984) data may be due to the fact that he analyzed twin data. The different results on the use of indefinite and definite articles may result from different levels of analysis. While Szagun (2004a) analyzed article use in a group of six children at different MLU levels and performed error analyses for the total speech sample of 44 hours' speech per child, Mills (1986) analyzed a much smaller data set per individual child. Thus, on a descriptive level, the question of how German-speaking children acquire gender marking has not yet received a satisfactory answer. Perhaps more importantly, the question of whether young children make use of the regularities in gender marking which exist in the language system has not been examined in any depth at all.

The aim of the present research is to investigate whether young children use phonological regularities of noun structure when they acquire noun gender. While phonological, morphological and semantic regularities may interact in gender assignment, it is argued here that for children, initially, the most important regularities are phonological patterns in word final and word initial position. This is because the sound patterns of their native language come to be meaningful cues to structural patterns even of a grammatical nature very early on in development. Research in speech perception has shown that during the latter half of their first year, infants become sensitive to the phonotactic patterns of words in their native language, particularly concerning word boundaries (Jusczyk, Luce & Charles-Luce, 1994; Jusczyk, Friederici, Wessels, Svenkerud & Juscyzk, 1993). At this age, too, German-learning infants are able to recognize unstressed closed-class lexical elements in fluent speech (Höhle & Weissenborn, 2003). As we have seen, gender assignment in German is associated with phonological regularities, in particular the regular organization of sounds in word final and word initial position, and gender is marked on closed-class elements such as determiners. If infants have learned something about the regularities of the organization of sounds in word final and word initial position and are able to recognize unstressed determiners, then these abilities are likely to help them in learning gender assignment in their second year.

449

We would argue that, initially, phonological patterns are most important in guiding children's acquisition of noun gender because such patterns can be organized meaningfully by the young children's information processing system in this developmental period. While some phonological patterns may also have morphological and semantic functions, semantic cues especially are unlikely to be important organizing principles in young children's gender assignment. This is because the types of semantic fields associated with gender in German involve fairly abstract superordinate categories (Köpcke & Zubin, 1984; Heidolph *et al.*, 1984), and young children do not have the cognitive prerequisites for abstracting such common features. Empirical evidence for the prevalence of phonological cues in young children comes from the French-speaking children studied by Karmiloff-Smith (1979). In a series of experiments with nonsense words, Karmiloff-Smith (1979) showed that from three years onwards gender assignment was predominantly based on phonological cues in word endings in French-learning children. She argues that as early as three years, French-speaking children have constructed a system of phonological regularities in word endings which are associated with gender assignment.

We hypothesize that young German-learning children use the probabilistic phonological patterns in word initial and word final position which are associated with gender assignment when they acquire noun gender. If this is so, children should commit fewer errors when gender assignment conforms to the regularities than when it does not. Furthermore, children's errors should be systematic. Thus, they are expected to overgeneralize *der* when erring on non-masculine monosyllabic nouns and overgeneralize *die* when erring on non-feminine polysyllabic nouns ending in -*e*. Besides phonological regularities, frequency of word type is expected to influence the acquisition of noun gender. The role of frequency seems unclear at present. There are suggestions that for frequent words belonging to children's 'core' vocabulary, article forms may be learned by rote in association with these nouns (Mills, 1986). It is hypothesized here that frequency may influence the acquisition of gender assignment to the extent that the noun gender of more frequent word types is learned faster and with fewer errors.

The present study also aims at clarifying some of the unsettled questions in the area of article acquisition on the descriptive level. These concern the age of onset of gender marked article forms on which the data of Mills (1986) and Clahsen (1984) diverge. This issue is by no means trivial, as current diagnostic instruments of German child language (Clahsen, 1986; Grimm & Doil, 2000) work on the assumption of late article acquisition. Further, it would seem that the reported correct and erroneous use of definite and indefinite articles needs to be examined on the basis of a larger data set. While both Mills (1986) and Szagun (2004a) report more frequent use of indefinite articles, results with respect to error rates in the two

450

paradigms are inconclusive. Mills (1986) reports higher error rates for indefinite than for definite articles, whereas Szagun (2004a) reports no difference for the nominative case, which is by far the most frequent case. However, the two analyses may not be comparable, as Mills (1986) was referring to age levels around two years, while Szagun's (2004a) analysis was based on the total number of errors between ages 1;4 and 3;8. Looking at error rates for definite and indefinite forms over separate age levels may settle this question.

METHOD

Design and participants

The data for the present analysis are from a large corpus of the spontaneous speech of 22 German-speaking children (see Szagun, 2004a, b). The data are longitudinal. There are two subsamples. Six children, 4 girls and 2 boys, were recorded every 5–6 weeks between the ages of 1;4 and 3;8. There are 22 speech samples per child. Additionally, 16 children were recorded every 4½ months between the ages of 1;4 and 2;10. There are 5 speech samples per child. For each subsample, two hours' spontaneous speech in a free play situation were audiorecorded per data point. The children had no diagnosed developmental delays and demonstrated age-appropriate object permanence knowledge at the start of data collection at age 1;4 (Sarimski, 1987). The children were growing up in monolingual environments and were resident in Oldenburg, northern Germany. They were recruited for the study from two daycare centres and a paediatrician's practice in Oldenburg.

Data collection and transcription

Data collection took place in a playroom at the Department of Psychology at the University of Oldenburg. The situation was free play with a parent and – most of the time – an investigator present and playing with the child. There was a set of toys such as: cars and garage, dolls, doll's house, zoo animals, farm animals, forest animals, picture books, puzzles, medical kit, ambulance and fire station. Digital auditory tape recording was carried out, using portable Sony DAT recorders and high-sensitive Sony or Aiwa microphones. At data points 1;4, 1;8, 2;1 and 2;5, video recordings were also made.

Everything spoken by the child and the first 500 parental utterances at data points 1;4, 1;8, 2;1 and 2;5 were transcribed using CHILDES (MacWhinney, 2000). An adaptation of rules for transcribing contracted speech and for coding morphosyntax was developed and used (Szagun, 1999, 2004b). Transcription was done from the DAT recordings. Eight transcribers who were trained extensively on using CHAT notations and transcription rules performed the transcriptions. Reliability checks on

transcription were calculated for 7·3% of the total transcripts for different pairs of transcribers. Percentage agreements were between 96% and 100%. Coding of morphosyntax was performed by three researchers. Articles were coded as part of the overall morphosyntactic coding. Reliability checks were done for 20% of the total transcripts for the overall morphosyntactic coding which does not include error coding. Cohen's kappa was used as a measure of reliability. Kappas were 0·94 for coders one and two, 0·98 for coders one and three, and 0·96 for coders two and three, indicating very good agreement between coders. CLAN programmes were used for calculating frequencies of correct and erroneous gender marked articles.

Categorization scheme for noun vocabulary according to phonological regularities of gender assignment

The children's noun vocabulary was categorized according to the phonological regularities of gender assignment. The analysis is based on 132 two-hourly speech samples from 6 children between 1;4 and 3;8, with 22 speech samples per individual child, and 75 two-hourly speech samples from 15 children between 1;4 and 2;10, with 5 speech samples per individual child. (One child had to be excluded from the subsample of 16 because she produced too few nouns per category). The total number of child utterances analysed was 87,983. This count excludes those utterances that were excluded from MLU counts, i.e. imitations, routines, vocalisations, partially incomprehensible utterances and one-word utterances consisting of 'yes/no' answers or fillers. The noun vocabulary summed over all the children is the basis of the analysis. The categories are presented and explained below. Following the Last-Member Principle (Köpcke & Zubin, 1984), compounds are placed in the appropriate category for the last member of the compound. Table 2 presents the main categories, including examples and absolute and relative numbers of types and tokens of nouns used with correct gender assignment. Percentages were calculated out the total of all nouns, excluding the rest category.

I *Polysyllabic nouns ending in* -e

  a) *Rule-conforming gender assignment*
     Polysyllabic nouns ending in -*e* tend to be feminine (Mills, 1986). Example: *die stufe* 'step'.
  b) *Not rule-conforming masculine and/or neuter gender assignment*
     Some nouns ending in -*e* are masculine or neuter. Examples: *der name* 'name', *das auge* 'eye'.

II *Monosyllabic nouns with consonants in initial and/or final position*

  Monosyllabic nouns tend to be masculine (Köpcke & Zubin, 1983, 1984). Köpcke & Zubin (1983) specified a number of specific noun combinations in

452

TABLE 2. *Categorization of children's noun vocabulary according to phonological regularities of gender assignment with type and token frequencies of nouns used with correct gender assignment*

Category based on phonological regularity	Examples with translation	Number of words in children's vocabulary[a]			
		Types	(%)	Tokens	(%)
I Polysyllabic nouns ending in *-e*					
a) rule-conforming: gender assignment feminine	*die kerze* 'candle'	262	(21·1)	2005	(17·1)
b) not rule-conforming: masculine and neuter words	*das auge* 'eye', *der löwe* 'lion'	22	(1·8)	303	(2·6)
II Monosyllabic nouns with final and/or initial consonants					
a) rule-conforming: gender assignment masculine	*der strumpf* 'stocking', *der fuss* 'foot', *der schnee* 'snow'	253	(20·3)	2536	(21·6)
b) not rule-conforming: feminine and neuter words	*die stadt* 'town', *das bein* 'leg', *das reh* 'deer'	167	(13·4)	2337	(19·9)
III Polysyllabic nouns ending in *-el, -en* and *er*					
a) rule-conforming: gender assignment masculine	*der nagel* 'nail', *der regen* ' rain', *der eimer* 'bucket'	240	(19·3)	1977	(16·8)
b) not rule-conforming: feminine and neuter words	*die nadel* 'needle', *das fohlen* 'foal', *die leiter* 'ladder'	67	(5·4)	485	(4·1)
IV Polysyllabic nouns with specific endings and deterministic gender assignment					
1) masculine words ending in *-or, -ig, -ling, -ist*	*der motor* 'motor', *der könig* 'king'[b]	12	(1·0)	106	(0·9)
2) feminine words ending in *-ung, -ei, -ie, -ik, -in*	*die heizung* 'heating', *die bäckerei* 'bakery'[b]	29	(2·3)	115	(1·0)
3) neuter words ending in *-chen, -lein, -fon,* and with prefix *ge-*, and nominalized verb infinitives	*das kindchen* 'little child', *das gesicht* 'face' *das telefon* 'telephone'[b]	78	(6·3)	385	(3·3)
V Polysyllabic nouns with no common regularities	*das auto* 'car', *die arbeit* 'work', 'elephant'	114	(9·2)	1448	(12·3)
	Total	1244		11747	

[a] The numbers are based on 132 two-hourly speech samples from 6 children and 75 two-hourly speech samples from 15 children.
[b] Examples for each ending are presented in the text.

word initial and/or word final position with varying probabilities of masculine gender assignment. However, their analysis is based on adult written vocabulary and had to be modified for a number of reasons for present purposes. In our database of the children's vocabulary the number of words per category as specified by Köpcke & Zubin was small. But there were many monosyllabic words with differing consonants and consonant clusters in initial and final position. Concerning the classes assigning feminine gender according to Köpcke & Zubin (1983), there were only very few words in the children's vocabulary. For words ending in -*ft* and -*cht*, only 12 types occurred, of which 9 were masculine, e.g. *der stift* 'pen', and 3 feminine, e.g. *die nacht* 'night'. Thus, on the basis of their productive vocabulary, children would be more likely to associate words ending in -*ft* and -*cht* with masculine gender. Concerning words ending in -*ur* and -*ür*, only 2 words, *die uhr* 'clock' and *die tür* 'door', occurred. This database would hardly be sufficient to discover a regularity. In view of our database and Mills' (1986) result that children did not associate these classes of words with feminine but rather with masculine gender, we subsumed words in these two classes in the category of monosyllabic words tending to be associated with masculine gender. Köpcke & Zubin (1984) themselves suggest that the five rules they specify for monosyllabic masculine nouns could be collapsed to one rule stating that the probability of a noun being masculine increases with an increasing numbers of consonants in word initial or final position. Following this suggestion, we use one main category for monosyllabic nouns encompassing monosyllabic nouns with consonants in initial and/or final position.

  a) *Rule-conforming masculine gender assignment*
     Monosyllabic nouns with one or more consonants in initial and/or final position tend to be masculine. Examples: *der tag* 'day', *der hund* 'dog', *der schrank* 'cupboard'.
  b) *Not rule-conforming feminine and neuter gender assignment*
     Other monosyllabic nouns with one or more consonants in initial and/or final position are feminine or neuter. Examples: *das bad* 'bath', *die maus* 'mouse', *die bank* 'bank, bench', *das wort* 'word'.

III *Polysyllabic words ending in* -el, -en, -er

   This category comprises monomorphemic nouns ending in -*el*, -*en* and -*er*. In standard grammatical treatments (Heidolph *et al.*, 1984; Duden, 1995) these words are treated as masculine words with no specification as to whether this gender association is of a probabilistic or deterministic nature. According to our database, it is probabilistic. The endings -*en* and -*er* can also have a morphological function, for instance -*en* marks the verb infinitive and nominalized verbs and -*er* can be used as a derivational suffix, as for instance in the noun *aufkleber* 'sticker'. In our categorization the latter was

subsumed in the general category of words ending in -*er*, as it is often hard to distinguish between a derived and a monomorphemic word when the word is not newly formed.

a) *Rule-conforming masculine gender assignment*
Polysyllabic nouns ending in -*el*, -*en*, and -*er* tend to be masculine. Examples: *der deckel* 'lid', *der wagen* 'wagon', *der käfer* 'beetle'.
b) *Not rule-conforming feminine and neuter gender assignment*
Some polysyllabic nouns in -*el*, -*en*, and -*er* are feminine and neuter: *die schaukel* 'swing', *das kabel* 'cable', *das kissen* 'cushion', *das fenster* 'window', *die butter* 'butter'.

IV *Polysyllabic words with specific endings and deterministic gender assignment*
In standard grammatical treatments a number of endings which are associated with a particular noun gender are specified (Heidolph *et al.*, 1984; Duden, 1995). Many of these are derivational morphemes, such as -*chen* and -*lein*, expressing diminutive, or -*in* referring to a feminine person. Here, only those classes of endings which occurred in the children's vocabulary are named. These endings are associated with a particular noun gender in a deterministic fashion. No exceptions are mentioned in the standard grammars, and there were none in our data, provided gender assignment was correct. Examples per gender are:

1) *masculine words ending in* -or, -ig, -ling, -ist
Examples: *der traktor* 'tractor', *der käfig* 'cage', *der schmetterling* 'butterfly', *der polizist* 'policeman'.
2) *feminine words ending in* -ung, -ei, -ie, -ik, -in
Examples: *die wohnung* 'flat', *die bücherei* 'library', *die batterie* 'battery', *die musik* 'music', *die freundin* 'female friend'.
3) *neuter words ending in* -chen, -lein, -fon, *and with prefix* ge- *and nominalized verbs*
Examples: *das entchen* 'little duck', *das vöglein* 'little bird', *das mikrofon* 'microphone', *das gewitter* 'thunderstorm', *das einkaufen* 'shopping'.

V *Polysyllabic nouns with no common regularities*
This category contains words which cannot be subsumed under any common regularities and their association to gender assignment. Examples are: *das auto* 'car', *der elefant* 'elephant', *die kamera* 'camera', *das bonbon* 'sweet', *das papier* 'paper'. Some of the words in this category are loan words. However, there are no regularities which would characterize a sizable portion of these words and their gender assignment.

These categories classified the complete vocabulary except some personal names and a few unclassifiable words, such as *t-shirt* and *pommes*. These comprised 27 names and 6 unclassifiable words, 33 types and 75 tokens. This amounted to 2·0% of word types and 0·6% of tokens. Deducting

455

'polysyllabic words with no common regularities' and the unclassifiable words from the total vocabulary, gender assignment for 88·8% of the children's noun types can be classified on the basis of different phonological and sometimes also morphological regularities of noun structure associated with gender.

Coding scheme for errors of gender

In the nominative of the definite paradigm the three article forms der_{MAS}, die_{FEM} and das_{NEU} can substitute for one another. In the indefinite paradigm $eine_{FEM}$ is substituted by $ein_{MAS, NEU}$, and *ein* by *eine*. Examples are: *der bein* ('leg', correct: *das bein*); *die bus* ('bus', correct: *der bus*); *das kaffee* ('coffee', correct: *der kaffee*); *ein flasche* ('bottle', correct: *eine flasche*); *eine ohr* ('ear', correct: *ein ohr*).

In the accusative the three definite forms den_{MAS}, die_{FEM}, das_{NEU} can substitute for one another, and in the indefinite paradigm the forms $einen_{MAS}$, $eine_{FEM}$, ein_{NEU}. Examples are: *für den auto* (correct: *das auto*, 'for the car'); *die stuhl dahin* (correct: *den stuhl*, '(put) the chair there'); *der will eine kind spielen* (correct: *ein kind*, 'he wants to play a child'); *has' du auch so ein brücke?* (correct: *eine brücke*, 'have you also got such a bridge?').

In the dative $dem_{MAS, NEU}$ and der_{FEM} can substitute one another, in the indefinite paradigm the forms $einem_{MAS, NEU}$ and $einer_{FEM}$. Errors involving gender and case also occur, such as accusative *den* substituting for feminine dative *der* and neuter dative *dem*, feminine *die* substituting for masculine and neuter dative *dem*, and neuter *das* substituting for feminine *der* and masculine *dem*. Examples are: *von den zahl* (correct: *von der zahl*, 'from the number'); *weil der in den buch is* (correct: *in dem buch*, 'because he is in the book'); *will mit die bauernhof spiel'n* (correct: *mit dem bauernhof*, 'want to play with the farm'). As hardly any dative forms of the indefinite article occurred, indefinite forms were not included in the analysis.

Correct and incorrect gender marked nouns of the vocabulary were categorized according to the coding scheme for phonological categories by an independent second coder using 64% of the speech samples. Coder reliability was 95%. Additionally, inter-rater reliabilities were calculated for error coding using 64% of the speech sample and Cohen's kappa as a measure of agreement. Kappas were: nominative definite article, 0·95, nominative indefinite article, 0·89, accusative definite article, 0·97, accusative indefinite article 0·93, dative definite article, 1·0. Disagreements between coders resulted from including an incorrectly pronounced word with incorrect gender marking in the word count or not. Examples are: *eine daffi* meaning *eine kaffee* ('coffee', correct: *ein kaffee*), *ein raff* meaning *ein giraffe* ('giraffe', correct: *eine giraffe*), *das schau* meaning *das zaun* ('fence' correct: *der zaun*).

456

RESULTS

For the present analysis gender marked definite and indefinite article forms were used irrespective of case. The most frequent forms were nominatives, 71·3% for correct and 64·5% for incorrect forms. Errors of case were ignored if gender was correctly marked. Articles used pronominally were not included in the analysis in order to exclude any possible uncertainties about reference. There are three sections of the analysis. First, a descriptive analysis of article use over age is presented. Next, correct and incorrect gender assignment in the different noun categories defined according to phonological regularities is explored. Finally, the relation of adult frequency of noun use and gender assignment by children is examined.

Descriptive analyses

This analysis is based on the longitudinal data from 6 children with 22 speech samples per child. Figure 1 presents absolute frequencies of gender marked definite and indefinite article forms, correct and incorrect forms summed up. There are similarities and differences between the 6 children. Initially, all the children used gender marked articles with very low frequency for a period of time. This was followed by a very sharp increase in article use (see Figure 1). Onset of article use occurred between 1;5 and 1;8 for 5 children, and at 2;1 for 1 child, EME. The length of time with infrequent use varied between children. For ANA it was the shortest, with only 4 months, and for RAH it was most extended, with 8 months. Regarding the initial onset of infrequent use there appeared to be no preference for indefinite or definite forms. FAL and SOE used indefinite forms earlier than definite forms, whereas EME and RAH used definite forms before indefinite forms, and ANA and LIS started with both paradigms simultaneously.

Next, we examined if the rapid increase in article forms occurred earlier for indefinite than for definite articles. For most children rapid increase occurred around a frequency of 45–50. Therefore, a frequency of at least 45–50 being reached for the first time was used as a cut-off point for frequent vs. infrequent use. The age at which children reached this frequency was compared for indefinite vs. definite articles. Mean ages were 2;4 for indefinite articles and 2;7 for definite articles. While frequent use occurred three months earlier for indefinite articles this difference was not significant (*t*-test). Thus, indefinite and definite articles are acquired around the same time, in terms of onset and subsequent frequent use. The data show that gender marked articles are acquired early by German-speaking children, being used infrequently from the middle of the second year and frequently by the middle of the third year. Definite and indefinite articles are used with varying frequencies over the data points.
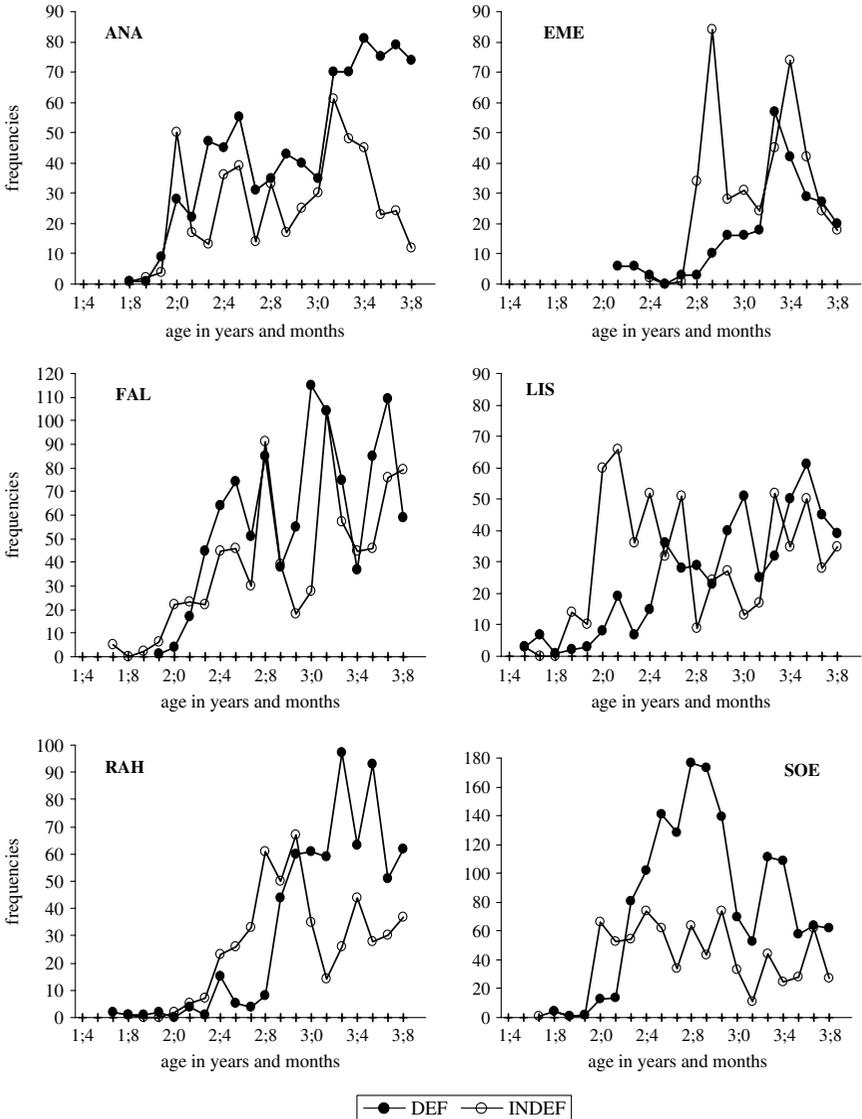
457

Figure 1. Frequencies of definite and indefinite gender marked articles (correct and erroneous forms) over age.

Figure 2 displays the percentage scores of correct definite and indefinite articles at the different age levels per individual child. Calculation of percentage scores was performed from the data point at which a minimum
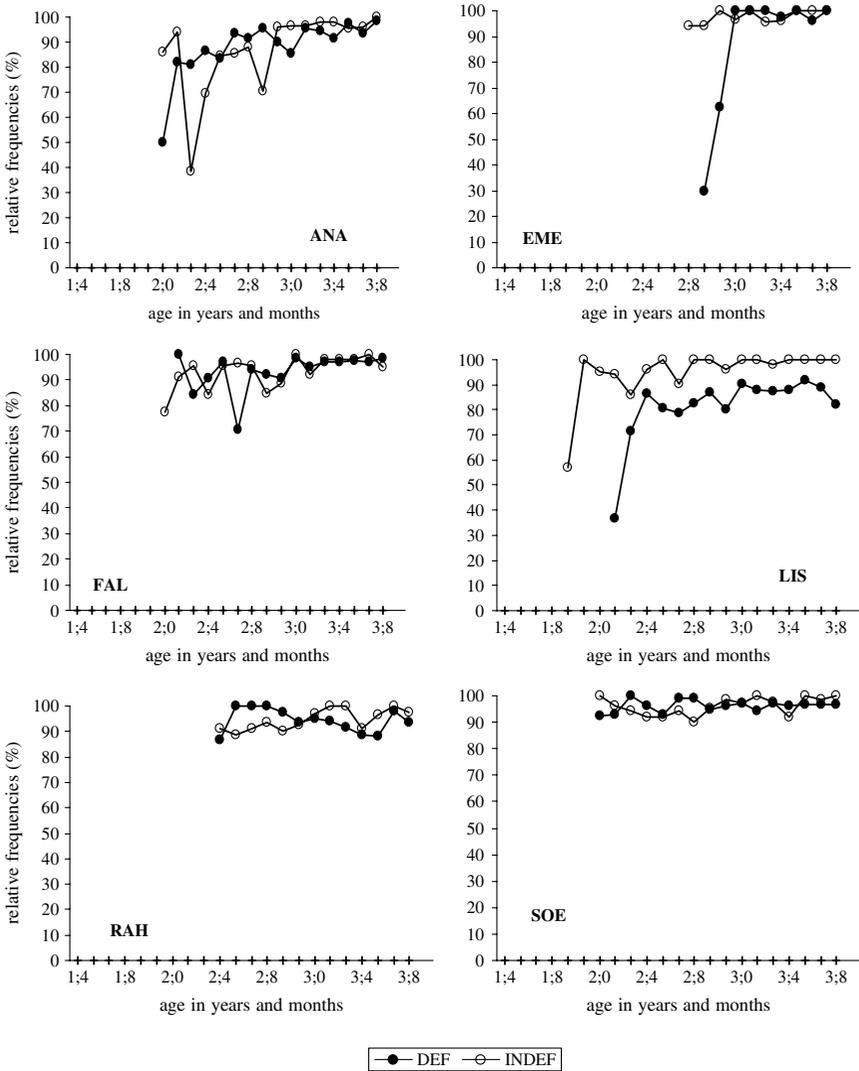
458

Figure 2. Percentage scores of correct forms of definite and indefinite gender marked articles over age.

of 10 forms, correct plus incorrect, was used. With one exception, all the children reached a level of 90% correct use for definite and indefinite forms during the period of data collection. For most of them this was around three years, but much earlier for SOE who used 90% correct forms almost from the beginning. LIS did not reach 90% correctness for the definite article.

459

The children varied considerably with respect to error rates. Particularly, ANA, FAL and LIS displayed high error rates for an extended period initially, EME did so for definite articles, but for a shorter period. Two comparisons between error rates for definite and indefinite forms were carried out, one comparing mean percentages of errors per article type over the whole age span, the other comparing mean percentages of errors over the initial time period before 90% correctness was reached. RAH and SOE, for whom no age trend is discernible, were excluded from the second comparison. For the whole age span mean ($n=6$) error scores were 7·6% for indefinite and 11·2% for definite forms. For the initial time period ($n=4$) error scores were 13·1% for indefinite and 25% for definite articles. These differences were not significant in either case (Wilcoxon's matched-pairs signed rank test).

Analysis of gender assignment to nouns in different categories
defined according to phonological regularities

Table 3 presents the absolute and relative frequencies (%) per noun category for children (correct use) and adults. Adult frequencies are based on 2,000 utterances per adult, collapsed over 21 adults. Percentages are calculated out of the total per category. As indicated in Table 3, there is remarkable agreement between child and adult use of noun types. In the category of polysyllabic nouns ending in *-e* the overwhelming majority of nouns is feminine. Regarding monosyllabic nouns, the tendency for these nouns to be masculine is already apparent in these young children's vocabulary. Masculine nouns comprise 60% and are twice as frequent as the next frequent category of neuter nouns. For polysyllabic nouns ending in *-el*, *-en* and *-er*, the percentage of masculine nouns is even larger. The majority of nouns with specific endings and deterministic gender assignment are neuter, and only a few are masculine. Adult and child use is very similar. In the category of polysyllabic nouns with no common regularities, for children the percentage of masculine and neuter nouns is similar, whereas feminine nouns are fewer. There is some discrepancy here between child and adult use, with adults using a somewhat larger proportion of feminine and a smaller proportion of neuter nouns. In all, however, even these young children use noun types per gender category with a distribution which resembles that in adult language very closely.

Comparison of errors of gender assignment in the different noun categories

For both subsamples of children, relative frequencies of errors of gender assignment to nouns in the different categories were calculated per noun category out of the total of correct and incorrect forms of definite and

TABLE 3. *Frequencies of masculine, feminine and neuter noun types per noun category as used by children* (*correct use*) *and adults*

| | Frequency of types (%) per gender | | | |
Phonological category	*Masculine*	*Feminine*	*Neuter*	Total
I Polysyllabic nouns ending in *-e*				
Children	18 (6·3)	262 (92·3)	4 (1·4)	284
Adults	10 (4·3)	224 (95·3)	1 (0·4)	235
II Monosyllabic nouns with initial/final consonants				
Children	253 (60·2)	41 (9·8)	126 (30·0)	420
Adults	219 (61·3)	44 (12·3)	94 (26·3)	357
III Polysyllabic nouns ending in *-el*, *-en*, *-er*				
Children	240 (78·2)	27 (8·8)	40 (13·0)	307
Adults	175 (74·2)	28 (11·9)	33 (13·9)	236
IV Polysyllabic nouns with specific endings and deterministic gender assignment				
1) masculine words ending in *-or, -ig, -ling, -ist*				
Children	12 (100)	0	0	12
Adults	9 (100)	0	0	9
2) feminine words ending in *-ung, -ei, -ie, -ik, -in*				
Children	0	29 (100)	0	29
Adults	0	29 (100)	0	29
3) neuter words ending in *-chen, -lein, -fon,* and with prefix *ge-*, and nominalized verb infinitives				
Children	0	0	78 (100)	78
Adults	0	0	87 (100)	87
V Polysyllabic nouns with no common regularity				
Children	49 (43·0)	18 (15·8)	47 (41·2)	114
Adults	43 (44·3)	25 (25·8)	29 (29·9)	97

indefinite articles. Calculations were done for noun types and noun tokens per category. Data were collapsed over age levels. For the subsample of 6 children this analysis is based on 22 speech samples per child, for the subsample of 15 children on 5 speech samples per child. Mean percentage scores of errors for all noun categories are presented in Table 4. In the subsample of 15 children, the number of noun types was insufficiently large to calculate percentage scores per child for some noun categories. Figures are presented for the subsample of 6 children only (see Table 4). For the categories 'polysyllabic nouns ending in *-e*', 'monosyllabic nouns with final and/or initial consonants' and 'polysyllabic nouns ending in *-el*, *-en* and *-er*' relative error frequencies in the rule-conforming category and the non-conforming category were compared.

461

TABLE 4. *Relative frequencies (%) of errors of gender assignment in the different noun categories*

Noun category based on phonological regularity	Mean frequency (%) of errors of gender assignment (SE)			
		$n=6$		$n=15$
	Tokens	Types	Tokens*	Types*
I Polysyllabic nouns ending in *-e*				
a) rule-conforming: feminine	7·5 (1·1)	15·5 (2·3)	9·5 (1·9)	16·3 (3·2)
b) not rule-conforming: masculine, neuter	4·8 (2·2)	15·1 (6·2)	—	—
II Monosyllabic nouns with final and/or initial consonants				
a) rule-conforming: masculine	4·6[a,e] (1·0)	9·1[b,g] (5·8)	6·4[c] (1·2)	11·0[d] (2·1)
b) not rule-conforming: feminine, neuter	11·4[a] (2·3)	22·6[b] (3·4)	15·7[c] (2·3)	19·8[d] (2·9)
III Polysyllabic nouns ending in *-el, -en, -er*				
a) rule-conforming: masculine	2·5[a,f] (0·6)	6·4[b,h] (1·5)	9·0[c] (3·4)	12·2 (3·5)
b) not rule-conforming: feminine, neuter	11·0[a] (2·9)	19·2[b] (3·6)	20·2[c] (2·9)	—
IV Polysyllabic nouns with specific endings and deterministic gender assignment				
	7·5 (7·3)	13·1 (12·6)	—	—
V Polysyllabic nouns with no common regularity	14·2[e,f] (15·0)	18·1[g,h] (3·3)	—	—

* Gaps indicate that there are too few data to calculate percentages.
[a, b, c, d, e, f, g, h] Means differ significantly (Wilcoxon, $p < 0.05$).

In the category 'polysyllabic nouns ending in *-e*' there was no significant difference in errors of gender assignment between the rule-conforming and non-conforming categories (Wilcoxon, n.s.). In the category 'monosyllabic nouns with initial and/or final consonants', children in both subsamples committed significantly more errors of gender assignment when nouns were not rule-conforming. This applied to noun tokens (Wilcoxon, $p < 0.05$ for the sample of 6, and Wilcoxon, $p < 0.05$ for the sample of 15) and noun types (Wilcoxon, $p < 0.05$ for the sample of 6, and Wilcoxon, $p < 0.03$ for the sample of 15). In the category 'polysyllabic nouns ending in *-el*, *-en* and *-er*', children in the subsample of 6 committed significantly more errors of gender assignment when nouns were not rule-conforming for tokens (Wilcoxon, $p < 0.03$) and types (Wilcoxon, $p < 0.05$). In the subsample of 15 children, calculations were performed for 10 children and for tokens only, because numbers were too low to calculate percentages. Errors of gender assignment were significantly larger when nouns were not rule-conforming (Wilcoxon, $p < 0.03$).

As Table 4 indicates, children also commit errors in the categories 'polysyllabic nouns with specific endings and deterministic gender assignment' and 'polysyllabic nouns with no common regularities'. Error rates in the three rule-conforming categories with probabilistic gender assignment were compared with error rates in the deterministic and the 'no rule' categories (see Table 4). For tokens and types, error rates were significantly lower in the categories 'monosyllabic nouns with initial and/or final consonants' and 'polysyllabic nouns ending in *-el*, *-en* and *-er*' than those in the 'no rule' category (Wilcoxon, $p < 0.05$). No other comparisons were significant. This indicates that a probabilistic rule of masculine, but not feminine, gender assignment leads to fewer errors than no rule. Deterministic gender assignment does not, however, lead to fewer errors than probabilistic gender assignment.

The next question is whether children err systematically, over-generalizing *der* when masculine gender is rule-conforming and *die* when feminine gender is rule-conforming. Only definite articles were used for this analysis, as only the definite paradigm has a distinct form for each gender. As there were insufficient data for definite article errors in the samples of 15 children, the analysis was carried out for the sample of 6 children only. The comparisons were carried out per non-conforming gender category in each phonological category. For monosyllabic feminine nouns, errors can be masculine or neuter. Masculine errors were significantly more frequent ($t(5) = 3.07$, $p < 0.03$). For monosyllabic neuter nouns, errors can be masculine or feminine. Masculine errors were significantly more frequent ($t(5) = 4.01$, $p < 0.01$). Means and standard errors are presented in Table 5. For polysyllabic nouns ending in *-el*, *-en* and *-er*, no error occurred for feminine nouns. For neuter nouns there was a non-significant tendency for

463

TABLE 5. *Mean frequencies of errors of gender assignment* (*tokens*) *for feminine and neuter non rule-conforming categories for the sample of* 6 *children*

Mean frequency of type of error (*SE*)	
Monosyllabic feminine nouns	
Masculine error	Neuter error
2·5[a] (0·8)	0·2[a] (0·2)
Monosyllabic neuter nouns	
Masculine error	Feminine error
18·7[b] (4·2)	2·5[b] (0·8)
Polysyllabic neuter nouns ending in -*el*, -*en* and -*er*	
Masculine error	Feminine error
1·7 (0·7)	0·2 (0·2)

[a, b] Means differ significantly, (*t*-test, $p < 0.03$).

masculine errors to be more frequent ($t(5) = 2.09$, $p < 0.09$). For non-feminine nouns ending in -*e*, frequencies of errors in the definite paradigm was so low that no comparison was performed.

There is the possibility that the masculine error preference is due to the fact that the majority of nouns in the vocabulary are masculine. Summing up the figures for noun types per gender category (see Tables 2 and 3) renders 572 masculine, 377 feminine and 295 neuter nouns in the children's vocabulary. Out of a total of 1,244 noun types, this means that 46·0% of all nouns are masculine, 30·3% feminine and 23·7% neuter. In order to control for a possible frequency effect favouring the masculine gender, it was checked to see if masculine error preferences exist for those noun categories which are not characterized by regularities associated with the masculine gender. This was done for feminine and neuter nouns in the categories with deterministic gender assignment, for polysyllabic nouns with no common regularities and for feminine nouns ending in -*e*. There was no significant preference for the masculine error in any of these categories.

Adult frequency of noun types and correct and incorrect gender assignment by children

The next question is whether adult frequencies per individual noun are related to frequencies of child correct and incorrect gender assignment to these same nouns. In accordance with previous vocabulary analyses (see Tables 2 and 3), the data from 21 children and adults were used. As a measure of capturing correct and incorrect child use of individual nouns, relative frequencies of tokens used with correct gender per individual noun were calculated out of the total of tokens with correct and incorrect gender assignment per noun. Only totals of 10 and larger were used for this

464

analysis, reducing the 532 noun types used by children and adults to 211. A Spearman rank correlation indicates that adult frequencies per noun type and children's relative frequencies of correct gender assignment per noun are significantly correlated (Spearman rho, $r = 0.14$, $p < 0.05$). Thus, the more frequently a noun is used by adults, the higher the frequency of correct gender assignment for that noun in child language, and vice versa. However, the percentage of variance explained by this correlation coefficient is only 1·96%.

DISCUSSION

On a descriptive level the present longitudinal results from six German-speaking children show that most children started using gender marked articles between 1;5 and 1;8. Onset of article use was around the same time for both indefinite and definite articles. After an initial period of infrequent use, frequency of articles increased sharply, for most children shortly after two years. Error rates were similar in both article paradigms and dropped well below 10% around the age of three years. There was considerable variation among the children concerning error rates.

The children's noun vocabulary was categorized according to phonological criteria of word structure associated with gender assignment. Monosyllabic words, and polysyllabic words ending in -el, -en and -er tend to be associated with masculine gender. Children committed more errors of gender assignment with nouns which did not conform to these regularities and fewer errors with nouns which did, thus confirming our hypothesis. For feminine nouns ending in -e, error rates did not differ in dependence on rule conformity or not. These results apply to both subsamples of children. Error preferences were tested in the subsample of six children. For neuter and feminine monosyllabic nouns, children overgeneralized *der* more frequently than *die* or *das*, thus erring systematically in the direction of the rule. For polysyllabic words ending in -el, -en and -er, there was a non-significant tendency in the same direction.

Adult frequencies per noun type were associated with relative frequencies of correct gender assignment per noun type in child language. Gender marking for more frequently used nouns was learned faster and was less error-prone.

Concerning descriptive details, the present data help to answer some of the open questions in the area of article acquisition. Firstly, our data show that German-speaking children start using gender marked articles very early indeed. They do so in the second half of their first year and thus even earlier than observed by Mills (1986) on the basis of early diary and observational studies. The sharp increase in article use shortly after two years is indicative of very rapid learning. Fast learning manifests itself in

465

rapidly decreasing error rates, too. The present data show that young German-speaking children have mastered the gender system by three years of age, when error rates have dropped well below 10%. This is astonishing, as error rates in other morphological paradigms, such as case and plural marking, are higher and persist well beyond three years (Clahsen, 1984; Mills, 1985; Tracy, 1986; Behrens, 2001; Szagun, 2001, 2004a). There is an insignificant trend in our data that indefinite articles are acquired earlier, confirming Mills' (1986) observation based on diary data. But the present results do not show higher error rates for indefinite articles, and thus do not confirm Mills' (1986) observation with respect to errors. What seems most characteristic of article acquisition is individual variability. Some children start with definite, others with indefinite articles, others with both. Some children go through an initial period of fairly high error rates, others do not. Only one child has consistently higher error rates in one paradigm, which is the definite article. Variability also characterizes which type of article, indefinite or definite, is used. Frequencies of articles in the two paradigms vary considerably over individual data points for every child (see Figure 1). This is probably an indication that choice of article type in a particular sampling situation is, at least partly, dependent on pragmatic factors such as the momentary play situation. It is much harder to interpret the individual variation in initial error rates. In so far as the tendency for higher or lower error rates is observed in other morphological paradigms for these same children (see Szagun, 2001), this could be due to different learning styles which are related to a more or less analytic style of learning (Bates, Bretherton & Snyder, 1988).

While the present descriptive results are limited to a sample of six children, this is still a larger sample than those in earlier studies (Clahsen, 1984; Mills, 1986), and the amount of speech sampled is considerably larger. Therefore, the results may be more generalizable. The discrepancy between observed age of acquisition in Mills' (1986) and Clahsen's (1984) data is resolved here in favour of early acquisition, thus supporting and even going beyond Mills' (1986) observations. This refers to onset of gender marked article use as well as level of correctness in gender marking which is indicative of acquisition, if one uses the conventional criterion of 90% correctness (Brown, 1973). This being the case, it would seem opportune that current diagnostic instruments of German child language which work on the assumption of late article acquisition (Clahsen, 1986; Grimm & Doil, 2000) adapt to this result.

Perhaps the more challenging question is why gender marking – a paradigm which is so difficult for second language learners – is acquired so rapidly by young children. One answer is that young children make use of the phonological patterns in word structure and their co-occurrence with gender marked articles. Such co-occurrences are of a probabilistic nature,

and young children's learning of gender assignment appears to be based on detecting such probabilistic regularities. Evidence for this is that children commit fewer errors when the regularities apply than when they do not. In the present data this evidence was strong for monosyllabic nouns and for polysyllabic nouns ending in -*el*, -*en* and -*er*, which tend to be masculine. Not only do children commit more errors when the regularities do not apply but they also err systematically in the direction of the rule assigning masculine gender to feminine and neuter nouns. One could argue that this error occurs because in the vocabulary as a whole the majority of nouns are masculine, and children just use the most frequent gender when they err. However, if the masculine error preference occurred on the basis of frequency alone, it should hold for errors in categories other than feminine and neuter monosyllabic nouns and polysyllabic nouns in -*el*, -*en* and -*er*. Thus, it should occur for feminine and neuter poly-syllabic nouns with no common phonological regularities, and for nouns with deterministic gender assignment, as well as for feminine nouns ending in -*e*. But there was no preference for the masculine error in any of these categories. Therefore, the overall higher frequency of masculine nouns in the language at large alone cannot explain the masculine error preference.

However, the argument for an influence of phonological regularities is weakened by the fact that children did not show a preference for assigning feminine gender to nouns ending in -*e*. There could be several reasons for this lack of a phonological regularities effect when feminine gender is required. One is that, due to the very low frequencies of noun types in the non-feminine categories (see Table 3), not enough definite article errors were sampled for the comparison with errors in the feminine category to be meaningful. Another possibility is that there is an effect of frequency. The few nouns which occurred in this category, such as *der junge* 'boy', *der hase* (literally 'hare' but used for 'rabbit'), *der löwe* 'lion' and *das auge* 'eye', were amongst the most frequently used words by children and adults. As highly frequent words tend to be used more frequently with correct gender assignment, this tendency could have overshadowed the effect of the phonological regularity. Another possibility is that for polysyllabic nouns ending in -*e*, the phonological regularities effect conflicts with the trend for assigning masculine gender on the basis of frequency. As a result, there is no effect in either direction. It is possible that phonological regularity effects exist on top of a general preference for masculine gender which is based on frequency. This could explain why they work for assigning masculine gender only. This interpretation would also seem to be supported by the fact that error rates for rule-conforming categories are lower than those for 'nouns with no common regularities' only when the rule-conforming categories are associated with masculine, but not with feminine, gender (see Table 4). We

467

conclude that our results demonstrate an effect of phonological regularities when these are associated with masculine gender.

The present analysis is the first to have demonstrated that even very young German-learning children make use of phonological regularities in noun structure when they learn gender assignment. We were able to show this for a sample of twenty-one children, irrespective of amount of speech sampled and age span covered. Some twenty years ago, Mills (1986) drew attention to the possibility of phonological regularities influencing children's acquisition of noun gender and was able to demonstrate that such regularities are operative in older children. The present results confirm Mills' (1986) findings and extend them beyond monosyllabic nouns. Together with the data from French-speaking children (Karmiloff-Smith, 1979), this is another set of data in a different language showing that young children acquire noun gender by exploiting the systematic relationship between the phonological structure of nouns and gender marked articles.

Besides phonological regularities, the frequency with which adults use individual nouns influences the acquisition of gender. Children learn gender assignment faster for nouns which are frequent in adult language. However, the influence of frequency is not strong, accounting for only 2% of the variance. A limitation of the present analysis is that all nouns used with a frequency of less than 10 (correct + incorrect use) had to be excluded from the analysis in order to have meaningful relative frequency counts. Thus, low frequency nouns in child language are excluded from the analysis. It could, however, be precisely the low frequency nouns which create difficulties in learning gender assignment for children. Despite this limitation, we can say that for high frequency nouns in child language, adult frequency has a mild influence on correctness of gender assignment.

The present analysis focuses on the influence of phonological regularities and their association with gender assignment as well as the influence of adult frequency on young children's acquisition of noun gender. This does not mean to say that other factors may not have an influence. Gender assignment based on semantic criteria, such as the common features underlying a semantic field, or natural gender, may well influence children's acquisition of gender marking. However, for most semantic fields, this influence is likely to become more important as children get older, because it presupposes world knowledge and cognitive abilities which are beyond the scope of the young children studied here. Learning gender assignment is also likely to be influenced by learning the case system. Case and gender are intertwined in German. While errors of case, errors of gender and errors of case and gender can be distinguished (see Szagun, 2004a), the two systems probably exert a mutual influence during acquisition. Thus, errors in the neuter paradigm which involve the *den/dem* substitution, i.e. *mit den buch* ('with the$_{ACC}$ book' – incorrect) instead of *mit dem buch* ('with the$_{DAT}$

468

book' – correct), may come about because children tend to substitute *den* and *dem* generally. They do so mostly in the masculine paradigm, when this constitutes an error of case only, as well as across gender paradigms when it constitutes an error of case and gender (Szagun, 2004a). The impact of this substitution for errors of gender studied here is very small, because the number of neuter nouns is small in the language. However, this type of error may come about because the acquisition of case and gender is, to some extent, intertwined. At the same time, the speed of acquisition of the gender system shows that a system which is only marginally affected by low discriminability of article forms, such as *den* and *dem*, is acquired much faster than a system which is characterized to a large extent by such forms (see Szagun, 2004a).

What is the significance of the present findings for theories of morphological learning? The present data add to the evidence for young children's sensitivity to patterns of regularity in the input language and thus support theories of statistical learning (Saffran, Aslin & Newport, 1996; Redington & Chater, 1998). Young children are sensitive to the recurrent phonological patterns in nouns, particularly in word initial and word final position. By the time children go about learning noun gender, they know about the regularities in the organization of sounds in their native language and they can attend to determiners (Jusczyk *et al.*, 1993, 1994; Höhle & Weissenborn, 2003). These are, of course, exactly the abilities which are needed for discovering the co-occurrence relationships between different phonological patterns in nouns and gender marked article forms. Thus equipped, young children exploit the information contained in the input about the probabilistic co-occurrence of gender marked articles and phonological patterns in noun structure. Language sounds are the type of information children can organize meaningfully in this developmental period. And they happen to be extremely helpful in learning noun gender assignment in German. Perhaps surprisingly, children do not treat highly consistent regularities differently from less consistent ones. They err no less with deterministic and near deterministic (nouns ending in -*e*) than with lower probability gender assignment. Maybe this in an indicator that deterministic rules are not different in kind from probabilistic ones, but just extreme cases of the latter (Altmann & Raettig, 1973).

It seems worth asking why noun gender is acquired so much faster than other morphological paradigms, such as plural marking, which may work along similar principles (Behrens, 2001; Szagun, 2001). This is probably due to the fact that combinations of gender marked articles and nouns are extremely frequent in the language. Thus, children's remarkable ability for statistical learning and the high frequency of occurrence of the phenomenon to be learned may conjointly work towards a fast acquisition process.

469

The acquisition of noun gender in German by children has remained a puzzle for a long time. How can children cope with a system which is so difficult to learn for adult second language learners? At least part of the answer is that children cope with this system because they attend to information which adult learners, perhaps, tend to ignore, and they rely more fully on statistical learning. It could be the specific learning mechanisms and types of information attended to at a particular time in development that make noun gender assignment not a difficult problem for young children.

REFERENCES

Altmann, G. & Raettig, V. (1973). Genus und Wortauslaut im Deutschen. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* **26**, 297–303.

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

Behrens, H. (2001). Learning multiple regularities: Evidence from overgeneralization errors in the German plural. In A. H.-J. Do, L. Dominguez & A. Johanse (eds), *Proceedings of the 26th Annual Boston University Conference on Language Development*, 72–83. Somerville, MA: Cascadilla Press.

Brown, R. (1973). *A first language: The early stages*. London: George Allen & Unwin.

Clahsen, H. (1984). Der Erwerb von Kasusmarkierungen in der deutschen Kindersprache. *Linguistische Berichte* **89**, 1–31.

Clahsen, H. (1986). *Die Profilanalyse: Ein linguistisches Verfahren für die Sprachdiagnose im Vorschulalter* (*Logotherapia 3*). Berlin: Marhold.

Duden, Band 4 (Volume 4) (1995). *Grammatik der deutschen Gegenwartssprache*, 5th ed. Mannheim: Dudenverlag.

Grimm, H. & Doil, H. ELFRA (2000). *Elternfragebögen für die Früherkennung von Risikokindern*. Göttingen: Hogrefe.

Heidolph, K. E., Flämig, W. & Motsch, W. (1984). *Grundzüge einer deutschen Grammatik*. Berlin: Akademie-Verlag, 2. Auflage.

Höhle, B. & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science* **6**, 122–7.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y. & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* **32**, 403–420.

Jusczyk, P., Luce, P. & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* **33**, 630–45.

Karmiloff-Smith, A. (1979). *A functional approach to child language*. Cambridge: Cambridge University Press.

Köpcke, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübingen: Niemeyer.

Köpcke, K.-M. & Zubin, D. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für germanistische Linguistik* **11**, 166–82.

Köpcke, K.-M. & Zubin, D. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte* **93**, 26–50.

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

MacWhinney, B., Leinbach, J., Taraban, R. & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language* **28**, 255–77.

Maratsos, M. (1982). The child's construction of grammatical categories. In E. Wanner & L. Gleitmann (eds), *Language acquisition: The state of the art*, 240–66. Cambridge: Cambridge University Press.

Mills, A. E. (1985). The acquisition of German. In D. Slobin (ed.), *The cross-linguistic study of language acquisition*, 141–254. Hillsdale, NJ: Erlbaum.

Mills, A. E. (1986). *The acquisition of gender: A study of English and German*. Berlin: Springer.

Redington, M. & Chater, N. (1998). Connectionist and statistical approaches to language acquisition. *Language and Cognitive Processes* **8**, 1–56.

Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-months-olds. *Science* **274**, 1926–8.

Sarimski, K. (1987). *Ordinalskalen zur sensomotorischen Entwicklung*. Weinheim: Beltz. (German version of Uzgiris, I. & Hunt, J. (1975). *Assessment in infancy: Ordinal scales of psychological development*. Champaign: University of Illinois Press.)

Szagun, G. (1999). *Rules for transcribing and analyzing German child language*. Institut für Kognitionsforschung, University of Oldenburg, Germany.

Szagun, G. (2001). Language acquisition in young German-speaking children with cochlear implants: Individual differences and implications for conceptions of a 'sensitive phase'. *Audiology & Neuro-Otology* **6**, 288–97.

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.

Tracy, R. (1986). The acquisition of case morphology in German. *Linguistics* **24**, 47–78.