Barbara Höhle, Jürgen Weissenborn, Dorothea Kiefer, Antje Schulz, Michaela Schmitz

Functional Elements in Infants’ Speech Processing: The Role of Determiners in the Syntactic Categorization of Lexical Elements

first published in:
Print ISSN 1525-0008
Online ISSN 1532-7078
DOI: 10.1207/s15327078in0503_5

Postprint published at the institutional repository of Potsdam University:
In: Postprints der Universität Potsdam : Humanwissenschaftliche Reihe ; 23
http://opus.kobv.de/ubp/volltexte/2008/1628/
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-16285

Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 23
How do children determine the syntactic category of novel words? In this article we present the results of 2 experiments that investigated whether German children between 12 and 16 months of age can use distributional knowledge that determiners precede nouns and subject pronouns precede verbs to syntactically categorize adjacent novel words. Evidence from the head-turn preference paradigm shows that, al-
though 12- to 13-month-olds cannot do this, 14- to 16-month-olds are able to use a
determiner to categorize a following novel word as a noun. In contrast, no categoriza-
tion effect was found for a novel word following a subject pronoun. To understand
this difference we analyzed adult child-directed speech. This analysis showed that
there are in fact stronger co-occurrence relations between determiners and nouns
than between subject pronouns and verbs. Thus, in German determiners may be more
reliable cues to the syntactic category of an adjacent novel word than are subject pro-
nouns. We propose that the capacity to syntactically categorize novel words, demon-
strated here for the first time in children this young, mediates between the recognition
of the specific morphosyntactic frame in which a novel word appears and the
word-to-world mapping that is needed to build up a semantic representation for the
novel word.

It is a common observation that most important morphosyntactic properties of the
native language seem to be acquired by the age of 3 years (cf. accounts for different
languages in Maratsos, 1998; Slobin, 1985). This observation immediately raises
the question of how this quick and relatively error-free learning is possible. Recent
hypotheses propose that elements of the functional vocabulary (i.e., inflectional
endings and function words) have an important role in how the child discovers lan-
guage-specific syntactic regularities of the native language. For example, in the
framework of the principles and parameters approach to language acquisition, it is
assumed that the information necessary for the determination of parametric values
is encoded in the functional vocabulary (Borer, 1984; Chomsky, 1981; Lebeaux,
2000). Within this framework, Hyams (1986) proposed that expletive pronouns
that have no referential meaning independent of their syntactic function as a
sentential subject may help the child to learn whether or not his or her target lan-
guage is a prodrop language that may leave out lexical subjects. Furthermore, it has
been pointed out that the fact that free and bound functional morphemes typically
occur at morphosyntactic boundaries might help the child to identify words and
syntactic phrases in the speech input (Christophe, Guasti, Nespòr, Dupoux, & van
Ooyen, 1997; Gerken, 2001; Morgan, Meier, & Newport, 1987; Shady & Gerken,
1999; Shafer, Shucard, Shucard, & Gerken, 1998). In addition, closed-class func-
tional morphemes may contribute to the syntactic categorization of content words
given the co-occurrence patterns between classes of functional morphemes and
classes of content words (e.g., Christophe et al., 1997; Gerken, 2001; Maratsos,
1982, 1988; Maratsos & Chalkley, 1980; Pinker, 1987; Shady & Gerken, 1999;
Shafer et al., 1998; Valian & Coulson, 1988).

There is converging evidence that infants are able to process unstressed func-
tional morphemes long before they start to produce them (Gerken, Landau, &
Remez, 1990; Gerken & McIntosh, 1993; Golinkoff, Hirsh-Pasek, & Schweiguth,
2001; Höhle & Weissenborn, 2003; Santelmann & Jusczyk, 1998; Shady &
Gerken, 1999; Shafer et al., 1998). Furthermore, children from 2 years on can use contextual cues given by functional elements to syntactically categorize nouns and adjectives (Eyer et al., 2002; Gelman & Markman, 1985; Taylor & Gelman, 1988). Similarly, from the age of 17 months on, the presence or absence of a determiner might lead the child to subcategorize a novel noun as a common noun versus a proper name (Gelman & Taylor, 1984; Katz, Baker, & Macnamara, 1974). Work with still younger children by Waxman and Booth (2001, 2003) shows that 14-month-old (but not 12-month-old) English-learning children can make use of the syntactic category of a novel word presented as a count noun (e.g., “a blicket”) or as an adjective (“blickish”), in construing the same set of objects as instantiating either a specific kind of object or a specific object property.

Our hypothesis is that such categorization may result from observation of the co-occurrence of the novel word with closed-class functional elements such as determiners. This categorization should go hand in hand with the imposition of specific co-occurrence restrictions on the novel word. If this is the case, we should be able to show that the child distinguishes between occurrences of the novel word in contexts that satisfy these co-occurrence restrictions and others that do not. We tested this hypothesis for determiners and subject pronouns with German infants of about the same age as the children in the Waxman and Booth (2001, 2003) studies, between 12 and 16 months of age, using the head-turn preference paradigm. Our prediction was that children would categorize novel words familiarized with preceding determiners as nouns and novel words familiarized with preceding subject pronouns as verbs. In our experiment, children were either familiarized with particular determiner–pseudoword sequences (e.g., ein glamm ‘a glamm’) or with particular personal subject pronoun–pseudoword sequences (e.g., sie glamm ‘she glamm’). After familiarization, all children were tested on passages in which the novel words were used in unambiguous noun phrase contexts as well as passages in which the pseudoword appeared in a potential verb context. If the type of familiarization leads to different categorizations of the novel words, infants of both familiarization conditions should react differently to text passages presented during testing, depending on whether the novel word occurred in a context compatible with the category it had been assigned to during the familiarization phase.

**EXPERIMENT 1**

**Methods**

*Participants.* Forty-eight infants between 12 and 13 months of age were tested. All children’s parents were native speakers of German. Participants were
assigned to one of two different familiarization conditions: determiner familiarization or pronoun familiarization (see later). The determiner familiarization group consisted of 24 infants—8 girls and 16 boys—with a mean age of 12 months 28 days (range = 12 months 2 days–13 months 30 days). The pronoun familiarization group also comprised 24 infants—12 girls and 12 boys—with a mean age of 12 months 21 days (range = 12 months 2 days–13 months 23 days). Data from an additional 22 infants were not included in the analysis of the results for the following reasons: incomplete experimental sessions (n = 18), average listening times under 3 sec in one of the experimental conditions (n = 3), or technical problems with the equipment (n = 1).

Materials. Two monosyllabic pseudowords (glamm, pronk) were selected. Monosyllables were used to exclude potential prosodic cues to the syntactic category of the items (cf. Kelly, 1996). Furthermore, the pseudowords did not contain any phonotactic cues that might make either a noun reading or a verb reading more probable. To avoid the necessity of an inflectional ending when used in a sentential context, the pseudowords were constructed in analogy to the third-person past tense form of the so-called strong verbs of German. The strong verbs—among which are very high-frequency German verbs—have a stem vowel change (Ablaut) but no inflectional ending in the third-person past tense form, for example, *sie schwimmt* (third-person present tense: ‘she swims’) and *sie schwamm* (third-person past tense: ‘she swam’). Because only a specific subset of vowels is involved in the Ablaut, for example, i → a (schwimmen—schwamm), e → o (fechten—focht ‘to fight’—‘fought’), only vowels of this subset were used in the pseudowords.

For the familiarization phase, the pseudowords were combined either with (a) the indefinite article *ein*—that is, *ein glamm* (‘a glamm’) and *ein pronk* (‘a pronk’)—turning the pseudoword into a possible masculine or neuter noun; or (b) the personal pronoun *sie*—that is, *sie glamm* (‘she glamm’) and *sie pronk* (‘she pronk’)—constituting a possible pronoun–verb sequence. For each pseudoword, two six-sentence passages were constructed in which the pseudoword occurred once in each sentence. (The material used in the experiments can be found at http://www.infancyarchives.com.) In one of these two passages, the pseudoword was consistently used in contexts that required a noun (noun passages), whereas in the other passage, the pseudoword occurred in contexts that required a verb (verb passages). None of the passages contained either of the function words used during the familiarization phase (i.e., *ein* or *sie*).

All stimuli were recorded by a German female speaker. She was instructed to read the material in a lively voice as if speaking to an infant. The stimuli were digitized (sampling frequency of 20,000 Hz, 16 bits mono) and transferred as sound files to a computer.
The children were familiarized either with determiner–pseudoword sequences (e.g., ein glamm), consisting of 26 tokens separated by pauses of 600 msec, or with pronoun–pseudoword sequences (e.g., sie glamm), consisting of 27 tokens, equally with pauses of 600 msec between them. The sequences had a duration of approximately 30 sec. The four text passages had a mean duration of 18.01 sec (noun passages: glamm = 18.63 sec, pronk = 17.21 sec; verb passages: glamm = 17.59 sec, pronk = 18.61 sec).

Procedure. For the experiment, a variation of the head-turn preference paradigm developed by Jusczyk and Aslin (1995) was used. In this method, an initial familiarization phase was followed directly by a test phase. During the experiment, the child was seated on the lap of a caregiver in the center of a test booth. The caregiver listened to music over headphones to prevent influences on the child’s behavior. Inside the booth, three lamps were fixed: a green one at the center wall and red ones at each of the side walls. Directly above the green lamp on the center wall was a hole for the lens of a video camera. On the outside of the test booth, two loudspeakers were mounted at the same height as the red lamps. Each experimental trial was started by the blinking of the green center lamp. When the child oriented on the green lamp, this lamp went out and one of the red lamps on a side wall started to blink. When the child turned his or her head toward the red lamp, the speech stimulus was presented from the loudspeaker on the same side as the blinking red lamp. The trial ended when the child turned his or her head away for more than 2 sec, or when the end of the speech file was reached. If the child turned away for less than 2 sec, the presentation of the speech file continued but the time spent looking away was not included in the total listening time.

The experiment was controlled by the experimenter from an adjacent room. She could start the visual and the acoustic stimuli by a push-button box that was connected to a computer. With the same means she coded the head turns of the child online during the experiment. From these codings the duration of the baby’s head turns during each experimental trial were calculated automatically. The experimenter observed the child’s behavior on a mute video monitor. Because the sound was not transmitted to the coding room, the experimenter was blind to the experimental condition to which a given trial belonged, so that coding behavior was not influenced by knowledge of the ongoing trial. The whole session was videotaped. In addition to the recorded picture from the testing booth, this videotape contained a code number for the baby being tested, a code number for each experimental trial, and a visual signal for the start of the acoustic stimulus. This additional information allowed for an offline coding of the infant’s orientation times to check the reliability of the online codings. Such an offline coding was done for 30% of the experimental sessions by a second rater. For the experiment reported here, the interrater correlation is $r = .93$, which is in the range reported by other studies us-
ing this method (Echols, Crowhurst, & Childers, 1997; Gomez & Gerken, 1999; Jusczyk, Hohne, & Bauman, 1999).

The children of the determiner familiarization condition were familiarized with the two determiner–pseudoword sequences, whereas the children of the pronoun familiarization condition were familiarized with the two pronoun–pseudoword sequences. The familiarization was stopped as soon as the children had listened to each of the sequences for at least 30 sec. During the test phase, all four passages were presented four times to each child. Thus, the pseudoword was used in a syntactic position incompatible with its use in the familiarization in half of the test passages. The passages were arranged in four blocks, each of which contained the four passages in a different order. The order of presentation of the four blocks was randomized between the infants.

Results and Discussion

The children familiarized with the determiner–pseudoword sequences had average listening times of 6.41 sec ($SD = 2.15$ sec) for the noun passages and 6.54 sec ($SD = 2.73$ sec) for the verb passages. A paired $t$ test revealed that the difference between the listening times was not significant, $t(23) = 0.35$, $p = .73$. Twelve of the children tested showed a preference for the noun passages. The mean listening times of the children familiarized with the pronoun–pseudoword sequences were 6.65 sec ($SD = 2.48$ sec) for the noun passages and 6.88 sec ($SD = 1.51$ sec) for the verb passages. Again, the difference proved not to be statistically significant, $t(23) = 0.44$, $p = .66$. In this condition, 13 infants had longer listening times to the noun passages.

These results show no systematic differences in the listening times to the passages used in the test phase of the experiment. Thus, there is no evidence that the structural context in which the pseudoword occurred during the familiarization phase led to a corresponding syntactic categorization of the pseudoword as either a noun or a verb. This could mean that, even if the child between 12 and 13 months of age has already formed a discrete representation of a determiner or a pronoun, he or she does not yet seem to have derived any co-occurrence restrictions for these items from the input that could then be transferred to the adjacent pseudowords. Our findings are consistent with the results of the study by Waxman and Booth (2003) who did not find a syntactic category effect on the 12-month-old child’s capacity to form different categories. This outcome should be predicted if—as is consistent with our results—the syntactic frame in which a novel word is used does not lead to syntactic categorization of the novel element. Correspondingly, given that the findings of Waxman and Booth (2001) with 14-month-olds presuppose the capacity of the child to categorize a novel word, we predict that, around the same age, the familiarization with the combination of
either a determiner or a subject pronoun and a following pseudoword as in the preceding experiment should now result in a categorization of the pseudoword as a noun or a verb. To test this, we ran the experiment with a group of 14- to 16-month-olds.

**EXPERIMENT 2**

**Methods**

*Participants.* Forty-eight infants between 14 and 16 months of age were tested. All children’s parents had German as their native language. Participants were assigned to the two different familiarization conditions: the determiner familiarization and the pronoun familiarization. The determiner familiarization group consisted of 24 infants—11 girls and 13 boys—with a mean age of 15 months 18 days (range = 14 months 4 days–16 months 24 days). The pronoun familiarization group also comprised 24 infants—again 11 girls and 13 boys—with a mean age of 15 months 21 days (range = 14 months 7 days–16 months 26 days). The data of 21 additional infants were not included in the analysis of the results for the following reasons: incomplete experimental sessions (n = 18), too short average listening times (n = 1), or technical problems with the equipment (n = 2).

*Stimuli and procedure.* The material and the procedure were the same as described for Experiment 1. In this experiment the interrater correlation for the coding of the infants’ looking behavior was \( r = .96 \).

**Results and Discussion**

The children familiarized with the determiner–pseudoword sequences had average listening times of 6.95 sec (SD = 2.30 sec) for the noun passages and 7.90 sec (SD = 2.85 sec) for the verb passages. A paired \( t \) test revealed that the difference between the listening times was significant, \( t(23) = 2.86, p < .01 \). Inspection of the individual data revealed that 17 of the infants had longer listening times for the verb passages. The mean listening times of the children familiarized with the pronoun–pseudoword sequences were 6.22 sec (SD = 1.97 sec) for the noun passages and 6.14 sec (SD = 2.53 sec) for the verb passages. This difference was not statistically significant, \( t(23) = 0.21, p = .83 \). In this condition, 14 of the infants listened longer to the verb passages.

These results indicate an effect of the familiarization for those infants familiarized with the determiner contexts. They prefer to listen to the verb passages. This novelty effect could be due to the occurrence of the pseudoword in a syntactic con-
text that is not expected by the children. This implies that they have classified the novel word as a noun by its co-occurrence with the determiner in the familiarization sequence. That this preference for the verb passages has to do with the familiarization and does not simply reflect a general preference for the verb passages is evidenced by the finding that the infants familiarized with the pronoun context do not prefer any of the passages.

One has to ask why we find a preference for the passages in which the pseudoword appears in categorically inconsistent contexts, leading to an ungrammatical sentence, given the fact that other studies found a preference for grammatical over ungrammatical sentences (Santelmann & Jusczyk, 1998). This parallels findings of studies with the same paradigm on infants’ word segmentation in which a preference to listen to familiar words (e.g., Jusczyk, Houston, & Newsome, 1999) as well as a preference to listen to unfamiliar words (Echols et al., 1997; Saffran, Aslin, & Newport, 1996) has been found. The outcome of either habituation or dishabituation effects is also known from experiments in the visual and the nonverbal acoustic domain—a fact that has been attributed to several factors including age of the infants tested, amount of familiarity with the stimuli, and type and complexity of the stimuli used (Burnham & Dodd, 1998; Hunter & Ames, 1988; Roder, Bushnell, & Sasseville, 2000; Schilling, 2000; Wagner & Sakovitz, 1986; Wetherford & Cohen, 1973). For experiments with speech stimuli, possible predictors for the outcome of a familiarity or a novelty effect in a given experiment have not been investigated so far but a proposal by Roder et al. (2000) can easily be transferred from the visual to the acoustic–verbal domain. They stated: “To the extent that the stimulus is congruent with something already represented in memory, attention to it is inhibited and deployed elsewhere” (p. 492). This proposal leads to a coherent explanation for the pattern of results found in our experiment. The children familiarized to the determiner context have built up a lexical representation for the novel word that includes the category information “noun.” When listening to the noun passages, the usage of the word in these passages is congruent with this lexical representation, leading to a loss of interest to these passages. In contrast, the usage of the novel word in the verb passages does not correspond to this memory representation, keeping the attention to this type of passages for a longer time on a higher level.

GENERAL DISCUSSION AND CONCLUSIONS

Summarizing the results of our experiments, the following picture emerges: 14- to 16-month-olds—but not 12- to 13-month-olds—show a differential effect of being familiarized to a novel word in either a determiner context or a subject pronoun context. These children showed longer listening times for the use of the novel word in a syntactic context different from the one in which the novel word had been fa-
miliarized if the familiarization context was constituted by a determiner but not when the context consisted of a pronoun. This suggests that the infants use the information given by the determiner to classify the following novel word as a noun but that they do not use a subject pronoun to classify the following word as a verb. These results are evidence that German 14- to 16-month-olds must already have acquired knowledge about syntactic co-occurrence restrictions between word classes, in our case between determiners and nouns, and that they can use this knowledge in turn to classify unknown words syntactically. In contrast, we did not find any evidence that the children had a similar knowledge of co-occurrence restrictions for subject pronouns that could be used to classify the following pseudoword as a verb.

One factor that might contribute to explaining this difference between the indefinite article and the personal subject pronoun could be that the evidence in the child’s linguistic input concerning the co-occurrence relation between the indefinite article and nouns is much stronger than the one concerning the co-occurrence relation between the personal subject pronoun and verbs. Thus, in German, due to its relatively free word order, a subject pronoun could be immediately followed by any of the following syntactic categories: a determiner, a noun, a prepositional phrase, or an adverb.

To verify our assumptions about the differences in the co-occurrence patterns of determiners and subject pronouns, we analyzed a corpus of child-directed speech. This corpus comprised 16 recordings of spontaneous utterances of a mother to her child during natural interactions. The recordings were made when the child was between 18 and 27 months old. The size of the corpus was approximately 15,000 words. Analyzing the occurrence of the determiner ein and the pronoun sie used in the familiarization condition of our experiments, we found that the indefinite article ein was used 179 times by the mother. In 71% of these occurrences, the article was immediately followed by a noun. The personal pronoun sie was used only 64 times. The pronoun was followed immediately by a verb in only 31% of these cases.

The results of this analysis suggest that, in German, a personal pronoun might be a less reliable cue to the syntactic class of the following word than the indefinite article. Thus, the reactions of the children in our experiment seem to reflect the distributional pattern for the determiner and the subject pronoun in their input. If the higher variation in the co-occurrence patterns of subject pronouns in German were the reason for their not being such a reliable cue to the syntactic category of an immediately following novel word, we would predict a different pattern in English where—given its almost unexceptional subject–verb word order—subject pronouns always precede a verb except in question contexts. This points to the interest of cross-linguistic comparative studies for testing specific developmental predictions.

Our finding that it is harder for the child to identify and to categorize verbs in the input than nouns is in accordance with the results of studies with older children
(Eyer et al., 2002; Olguin & Tomasello, 1993; Tomasello & Olguin, 1993). Furthermore, computational simulations of syntactic classification point to the same direction. Durieux and Gillis (2001) found that verbs are less reliably categorized as verbs than nouns as nouns by using phonological cues of the word form, such as stress pattern, segment quality, and length. Mintz, Newport, and Bever (2002) reported better success rates for the classification of nouns than of verbs when using the distributional contexts of the words to be classified. These observations suggest that—at least in the languages used in these simulations, namely English and Dutch, which are structurally very similar to German—there seem to exist less reliable input cues for the syntactic classification of verbs than for nouns.

Taken together, our results have provided the first clear evidence that, around 14 to 16 months of age, children must have built up lexical representations for determiners containing information about the language-specific syntactic co-occurrence restrictions of these elements and that this knowledge can be used by the child to assign immediately adjacent novel words to a syntactic category defined by these specific co-occurrence restrictions. Furthermore, our results show that children of this age are able to project the category information derived from one specific syntactic frame in which a novel word appears into other frames of the same syntactic class. This provides evidence for the existence of abstract category knowledge at least for nouns. Given the age of the children and the structure of the test passages on the basis of which we obtained our results, we also suggest that this knowledge initially is formal (i.e., not meaning based in nature). This is further support for the claim made by Naigles (2002) about children’s early abstract linguistic knowledge.

It is still an open question when these early formal representations of determiners are completed by information about their semantic properties such as definiteness. What we have to say is that the existence of representations of determiners in the child’s mental lexicon is a necessary prerequisite for acquiring these properties. These, in turn, allow the child to make further important distinctions in the syntactic category of co-occurring elements such as the one between common and proper nouns observed by Katz, Baker, and Macnamara (1974) in 17-month-old English children. How exactly the acquisition of the semantic properties of determiners proceeds has to be the objective of future studies.

ACKNOWLEDGMENTS

This work was supported by Deutsche Forschungsgemeinschaft Grants HO 1960/5–1 and HO 1960/5–2 to Barbara Höhle and Jürgen Weissenborn. We thank Susan Powers for her help with the corpus analysis as well as Christiane Bohn and our students for their help in preparing and running the experiments.
REFERENCES


