



Cognitive Science 35 (2011) 1190–1205
Copyright © 2011 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2011.01187.x

“Frequent Frames” in German Child-Directed Speech: A Limited Cue to Grammatical Categories

Barbara Stumper,^a Colin Bannard,^b Elena Lieven,^{a,c} Michael Tomasello^a

^a*Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology*

^b*Department of Linguistics, University of Texas at Austin*

^c*School of Psychological Sciences, University of Manchester*

Received 18 September 2009; received in revised form 14 March 2011; accepted 11 May 2011

Abstract

Mintz (2003) found that in English child-directed speech, frequently occurring frames formed by linking the preceding (A) and succeeding (B) word (A_x_B) could accurately predict the syntactic category of the intervening word (x). This has been successfully extended to French (Chemla, Mintz, Bernal, & Christophe, 2009). In this paper, we show that, as for Dutch (Erkelens, 2009), frequent frames in German do not enable such accurate lexical categorization. This can be explained by the characteristics of German including a less restricted word order compared to English or French and the frequent use of some forms as both determiner and pronoun in colloquial German. Finally, we explore the relationship between the accuracy of frames and their potential utility and find that even some of those frames showing high token-based accuracy are of limited value because they are in fact set phrases with little or no variability in the slot position.

Keywords: Language acquisition; Child-directed speech; Corpus analysis; Distributional analysis; Grammatical categories; German

1. Introduction

A number of studies have suggested that children are skilled statistical learners who can find distributional regularities in the speech signal that may aid in a number of acquisition tasks (Morgan & Demuth, 1996; Saffran, Aslin, & Newport, 1996). The availability of large corpora and new computational techniques have made it possible to study which statistical

Correspondence should be sent to Barbara Stumper, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. E-mail: barbara_stumper@eva.mpg.de

regularities are both present in the language input and potentially exploitable. One widely studied acquisition task is how children might discern lexical categories, for example, *noun* and *verb*, from the utterances they hear. Co-occurrence environments of words have been assessed in several studies as one potential cue to the category of a word (Chemla et al., 2009; Mintz, 2003; Mintz, Newport, & Bever, 2002; Redington, Chater, & Finch, 1998). The aim of these analyses was not to model the actual procedures a child might use, but rather to examine one kind of information that is available in input.

In Mintz (2003), the distributional patterns investigated were *frequent frames*. A frequent frame is defined as a window of two context words (*A* and *B*), which frequently co-occur in a corpus with exactly one intervening target word (*A_x_B*). Mintz (2003) demonstrated that the 45 most frequently occurring frames in the six English child-directed speech corpora he investigated contained words in the *x* position that belonged almost always to the same category (mean categorization accuracy 91%–98%). Moreover, there was considerable consistency in the frames across corpora suggesting that this categorization mechanism provides robust information. Further, the non-adjacency of the two context words seems to be crucial for the mechanism to work. The categories derived from frequent frames are more reliable than those derived from the co-occurrence of two adjacent words, for example, *A_B_x* or *x_A_B* (Chemla et al., 2009). Along with results from behavioral and computational modeling studies showing that both adults and children and, under certain conditions, neural networks can utilize such non-adjacent dependencies (Freudenthal, Pine, & Gobet, 2008; Gómez, 2002; Gómez & Maye, 2005; Mintz, 2006; Monaghan & Christiansen, 2004), these findings suggest that frames as defined in Mintz (2003) could be the basis for children's initial lexical categories.

An important question is whether this high degree of predictability is restricted to English or might be found in other languages as well. The high accuracy scores of the frame-based analyses in English might be due to the fact that English has a relatively fixed word order and a morphologically simple system of function words. Chemla et al. (2009) successfully extended the frequent frames account to French. When using the same frequency threshold as in Mintz (2003) to select a set of frames they found that each frame exclusively contained words from only one category (categorization accuracy 100%). Their analysis was carried out over one small French corpus of child-directed speech.

However, Erkelens (2009) applied the frequent frames account to Dutch, by analyzing four relatively dense corpora of Dutch child-directed speech. Mean accuracy scores ranged between 40% and 71% indicating high variability of informational value over frames. Further, scores were significantly higher for token than for type accuracy. A perceptual study with 12- and 16-month-old infants learning Dutch that replicated Mintz (2006) found no evidence of categorization from frequent word frames thus challenging the cross-linguistic applicability of the distributional pattern. One question raised by the work of Erkelens (2009) is why frames in Dutch collected such inconsistent sets of grammatical categories. As Dutch and German are similar in some ways, a qualitative analysis of the German frequent frames will shed light on this.

The present study presents a similar frequent frames analysis for German. By comparison to both English and French, German has a less restricted word order. By comparison to all

three previously studied languages, German has a morphologically more complex determiner system since determiners are marked for case, number, and gender. Furthermore, in colloquial German, determiners are often used pronominally. It is therefore possible that frequent frames in German might be less accurate in categorizing words than for English (Mintz, 2003), French (Chemla et al., 2009), and Dutch (Erkelens, 2009).

2. Method

2.1. Child-directed speech corpus

Our analysis was carried out over a longitudinal corpus of German child-directed speech to a boy, who we refer to here as Leo. Leo's caregivers have higher education and speak Standard German. This is the largest sample of child-directed speech that exists for German. The present analysis is based on 58 one-hour recordings made between the ages of 2;0 and 2;2. All words in this part of the corpus had been automatically labeled for their grammatical category with a German version of the CHILDES MOR-program (Behrens, 2000; MacWhinney, 2000). All cases in which the program provided several possible grammatical categories were manually checked and then disambiguated. Before the distributional analysis was performed, all utterances that contain unintelligible speech were excluded. Further, all special CHILDES transcription postcodes (e.g., [+ I]), phonological fragments, pauses and interjections were removed.

2.2. Distributional analysis procedure

All frames in the input speech were counted and tallied for frequency of occurrence. Following Mintz (2003), utterance boundaries were not treated as framing elements, nor could frames cross utterance boundaries. For each frame, all intervening words (types and tokens) labeled with their grammatical category were stored together in a group. A total of 30,601 utterances composed of 154,523 word tokens and 5,158 word types were analyzed. Next, as in Mintz (2003), the 45 most frequent frames, which were a subset of all frames extracted by the procedure, were selected for further analysis. On average, 143 word tokens (range: 87–410) and 29 word types (range: 3–62) per frame were analyzed. Frames contained words from a range of categories, particularly (in descending order) main verbs, pronouns, adverbs, auxiliary verbs, and nouns.

2.3. Quantitative measure of categorization success

For each frame, we evaluated how well the distributionally defined categories correspond to the syntactic categories. This was done by calculating accuracy for each frame. To this end, all possible pairs of word tokens as well as word types in each frame were compared. A Hit was recorded when two items were from the same grammatical category, and a False Alarm was recorded when two items were from different grammatical categories. *Frame*

accuracy measures the proportion of Hits to the number of Hits plus False Alarms (Mintz, 2003). Next, in order to assess the degree to which words from the same category were found in the same frame, accuracy for each syntactic category was computed (Mintz, 2003, “completeness,” p. 97; we henceforth refer to this as *category accuracy*). In this case, a Hit was recorded when two words from the same category ended up in the same frame. A Miss was recorded when two words from the same category ended up in different frames. *Category accuracy* measures the proportion of Hits to the number of Hits plus Misses.

Following Mintz (2003), two different methods of categorization were used. In Standard Labeling, nouns and pronouns were grouped, as were main verbs and auxiliary verbs. In Expanded Labeling, all four were treated as distinct categories. In Standard Labeling, on average, 920 tokens (range: 24–3,998) and 187 types (range: 20–711) per syntactic category contributed to the analysis. In Expanded Labeling, on average, 715 tokens (range: 24–3,366) and 146 types (range: 20–585) per syntactic category contributed to the analysis.

As can be seen from Table 1, the amount of data used in the present analysis exceeds the average amount of data used by Mintz (2003) and Erkelens (2009) as well as that used by Chemla et al. (2009). The tokens collected by the frames accounted for 4% of the corpus. Moreover, the categorized types covered 66% of the tokens in the whole corpus indicating that the frames collect highly frequent words.

2.4. Comparison to chance categorization

To ensure that accuracy measures for frames and categories were significantly different from chance, all gathered word tokens as well as all gathered word types were randomly assigned to all 45 frames. By doing so, the category structure of the corpus (number and size of categories) was held constant, whereas its distributional structure was overridden. Ten thousand trials of this random distribution of word tokens and word types to frames were undertaken. The distributional structure found in the present analysis was one of the trials. Next, token and type accuracy for all frames and categories was computed out of each of these trials. The original and the random accuracy scores were then compared. The *p*-value for each frame and each category was calculated as the proportion of random accuracy scores that are greater than or equal to the accuracy in the real data set. We used Fisher’s Omnibus test (Haccou & Meelis, 1992) to combine all these *p*-values into a single measure

Table 1

Number of analyzed corpora, number of utterances, number of tokens and types found in the selected frames, percentage of corpus (tokens) accounted for by categorized types, and percentage of corpus (tokens) analyzed

Child	No. Corpora Analyzed	No. Utterances	Tokens Gathered in Frames	Types Gathered in Frames	% of Corpus Accounted for	% of Corpus (Tokens) Analyzed
Mintz (2003) (Means)	6	17,199	4,517	446	50	6
Chemla et al. (2009)	1	2,006	172	99	12	0.9
Analysis	1	30,601	6,450	564	66.3	4.2

of overall significance. We tested the individual p -values using the False Discovery Rate procedure (Benjamini & Hochberg, 1995) to see which accuracy scores were not significantly different from chance.

3. Results

3.1. Frame accuracy

Mean token accuracy for Standard and Expanded Labeling was .77 ($SD = .20$) and .64 ($SD = .19$), respectively (Table 2). Mean type accuracy for Standard and Expanded Labeling was .57 ($SD = .25$) and .42 ($SD = .20$), respectively. All accuracy scores were significantly higher than random (Fisher's Omnibus test, $p < .001$). Two conclusions follow: first, the frames do gather some relatively reliable evidence of categories but secondly, there is considerable variability within the frames as indicated by the relatively low accuracy scores as compared to English or French. The scores are similar to the Dutch accuracy scores.

The difference between token accuracy scores for Standard and Expanded Labeling was significant ($t(44) = 5.97$, $p < .001$), so was the difference between type accuracy scores for both labeling protocols ($t(44) = 8.821$, $p < .001$). This is due to the fact that there are frames that collect both main verbs and auxiliary verbs. By contrast, nouns and pronouns only rarely end up in the same frame. Furthermore, the differences between token and type accuracy scores were significant for both labeling protocols (token-type accuracy scores Standard Labeling: $t(44) = 6.763$, $p < .001$; token-type accuracy scores Expanded Labeling: $t(44) = 5.792$, $p < .001$). Hence, there are frames containing types of various syntactic categories, whereas token frequencies are skewed to favor one syntactic category.

Following this up, we investigated how evenly types are distributed within each frame. This was done by calculating normalized Shannon–Weaver values as a measure of diversity according to the formula $SW(X) = -\sum_{x \in X} p(x) \ln p(x) / \ln(\text{nr of items})$ (Zar, 1999). X is the slot position of a given frame, each x is a word that appears in that frame and $p(x)$ is the probability of seeing each x in that position. The values were then standardized (cf. denominator) resulting in scores ranging from 0 to 1. For example, a frame might gather

Table 2

Mean frame accuracy (SD) for Standard and Expanded Labeling (token and type) including mean accuracy (SD) of random categories

	Token Accuracy		Type Accuracy	
	Analysis	Random	Analysis	Random
Standard Labeling	.77 ^{a,e,g} (.20)	.45 ^a (.000)	.57 ^{b,f,g} (.25)	.37 ^b (.001)
Expanded Labeling	.64 ^{c,e,h} (.19)	.33 ^c (.000)	.42 ^{d,f,h} (.20)	.26 ^d (.001)

Note: ^{a,b,c,d}Scores differ significantly (Fisher's Omnibus test, $p < .001$).

^{e,f,g,h}Means differ significantly (paired t tests, $p < .001$).

ten different words. If one of these words has many more tokens than all the other words, the distribution of words is skewed and this frame gets a low Shannon–Weaver value. If all ten different words occur with almost equal frequencies, the frame gets a high Shannon–Weaver value. Thus, low values indicate low diversity; high values indicate high diversity. As we were interested in detecting each frame’s lexical specificity, verb inflections were taken into account such that, for instance, the different forms of the verb *machen* (to make), for example, *macht* (3rd Sg) or *machen* (3rd Pl) were treated as different word types.

The Shannon–Weaver values for the frames ranged between .18 and .95 with a mean of .65 for Standard Labeling and a mean of .66 for Expanded Labeling. Further, correlations between the Shannon–Weaver values and token accuracy scores were calculated. Scores were negatively correlated, $r_p = -.358$, $p < .05$ (Standard Labeling) and $r = -.628$, $p < .001$ (Expanded Labeling), respectively. The negative correlation plays out as follows. Roughly speaking, there are three kinds of frames (Fig. 1). First, there are frames that categorize a variety of different words while most or even all words belong to the same category and occur equally frequently. For these frames, both accuracy and diversity score within or above the 1 *SD* range around the mean. We chose the results from the Standard Labeling procedure as baseline to which all the other accuracy scores were then compared. Fig. 2 gives an example for such an accurate and diverse frame: *die_x_ist*¹ (accuracy: .83, diversity: .95). Second, there are frames showing high accuracy together with low diversity, for example, *ist_x_das* (accuracy: .84, diversity: .28, Fig. 3). One single word clearly dominates the slot position accounting for the lion’s share of all tokens (mean 85%, range: 74–94%). In the case of the above-mentioned frame the adverb *denn* (particle, no translation) covers 85% of all tokens. The adverb is phonetically reduced in most cases (*nn* instead of *denn*) indicating that the co-occurring lexical items are frequently repeated (Bybee, 2007; “Reducing Effect,” chap. 12). Third, there are frames that collect words belonging to a variety of categories while there is more than one word covering the majority of the tokens. Consequently, accuracy is low but diversity is still relatively high, for example, *ist_x_der* (accuracy: .30, diversity: .78; Fig. 4). Accuracy scores were not significantly different from

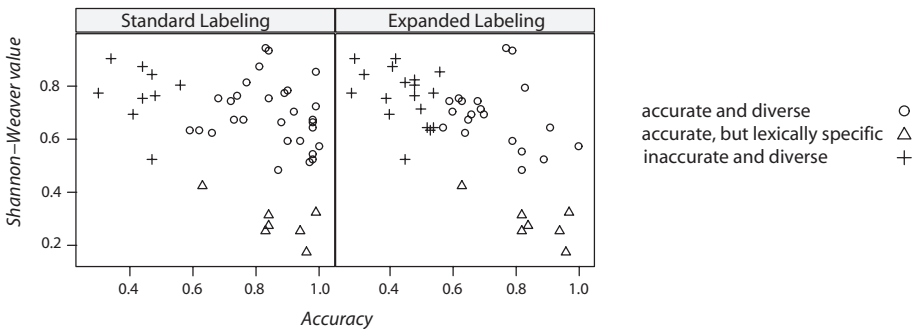


Fig. 1. Diversity (Shannon–Weaver values) as a function of accuracy: A_x_B frames.

random in five (Expanded Labeling) and eight (Standard Labeling) cases, respectively. As accuracy scores for Standard and Expanded Labeling differed significantly, the means for the three groups differ, respectively (Table 3).

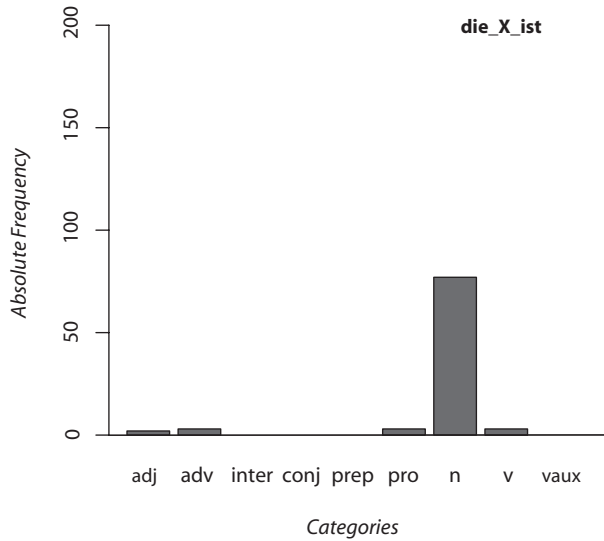


Fig. 2. Word token frequencies across categories (**adjective**, **adverb**, **interjection**, **conjunction**, **preposition**, **pronoun**, **noun**, **verb**, verb **auxiliary**) for an accurate and diverse frame *die__x__ist* (acc: .83, SW: .95; Standard Labeling).

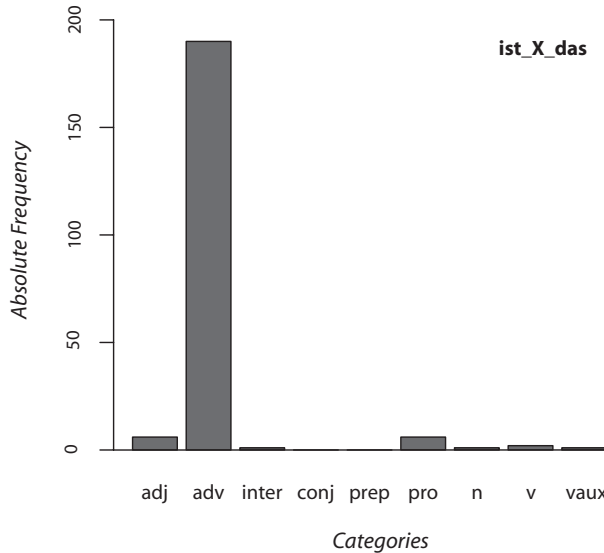


Fig. 3. Word token frequencies across categories (**adjective**, **adverb**, **interjection**, **conjunction**, **preposition**, **pronoun**, **noun**, **verb**, verb **auxiliary**) for an accurate but lexically specific frame: *ist__x__das* (acc: .84, SW: .28; Standard Labeling).

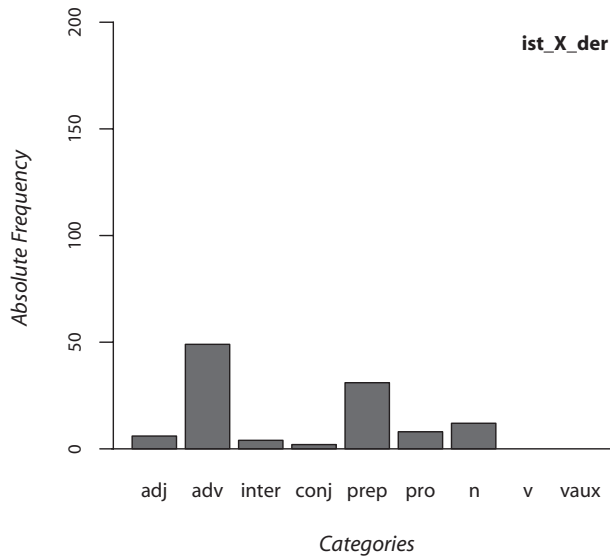


Fig. 4. Word token frequencies across categories (**adjective**, **adverb**, **interjection**, **conjunction**, **preposition**, **pronoun**, **noun**, **verb**, verb **auxiliary**) for an inaccurate frame: *ist_x_der* (acc: .30, SW: .78; Standard Labeling).

Table 3

Mean token accuracy (SD) and diversity (Shannon-Weaver) for three groups of $A \times B$ frames: Standard Labeling and Expanded Labeling

$A \times B$ Group	<i>n</i>	Standard Labeling		<i>n</i>	Expanded Labeling	
		Token Accuracy	Shannon-Weaver		Token Accuracy	Shannon-Weaver
Accurate and diverse*	30	.85 (.13)	.69 (.13)	20	.73 (.12)	.70 (.12)
Accurate but lexically specific*	6	.90 (.07)	.27 (.06)	7	.85 (.12)	.29 (.08)
Inaccurate and diverse*	9	.43 (.08)	.78 (.11)	18	.45 (.09)	.77 (.10)

Note: *‘Accurate’ corresponds to scores equal to or greater than .57. ‘Inaccurate’ corresponds to scores less than .57. ‘Diverse’ corresponds to scores that fall within or above the 1 *SD*-range around the mean. ‘Lexically specific’ corresponds to scores that fall below 1 *SD* below the mean.

Overall and as expected, frames rely heavily on closed-class words as framing elements (cf. Appendix).

3.2. Category accuracy

Mean token accuracy for categories was .10 (*SD* = .06) for Standard Labeling and .11 (*SD* = .03) for Expanded Labeling (Table 4). Mean type accuracy for categories was even lower, with .11 (*SD* = .07) and .07 (*SD* = .06), for Standard and Expanded Labeling, respectively. However, all scores were significantly different from chance categorization (Fisher’s

Omnibus test, $p < .001$). Words from the same grammatical category tended to occur in many different frames. The distribution of words of the same category across frames seems to be rather arbitrary. Not surprisingly and as found for English (Mintz, 2003), Dutch (Erkelen, 2009), and French (Chemla et al., 2009), one single frame does not offer a reliable cue to one category.

3.3. Partial frames

Perhaps, given the higher word order variability in German, a frame only involving A_x would work best. However, accuracy scores for A_x frames (Table 5) were considerably lower as compared to those derived from A_x_B frames.² The majority of A_x partial frames is inaccurate and diverse when compared to A_x_B frames (Table 6 and Fig. 5). As far as category accuracy is concerned, the tokens gathered by A_x frames account for 96.7% of the corpus compared to 66.3% for A_x_B frames.

4. Discussion

This analysis of German child-directed speech aimed to investigate whether, and how accurately, the category of a word could be derived from distributional information. The distributional information analyzed here was the use of *frequent frames* as defined by Mintz (2003). The results show that frames collected highly frequent words that covered 66% of all tokens in the corpus. As in English, Dutch, and French, German frames rely on function words as framing elements. But there was more variability within the frames. Nevertheless, all scores were significantly different from chance categorization indicating that frames could provide useful information. However, frames collected words more accurately at the level of tokens than at the level of types. Following this up, we found a clear negative correlation between accuracy of frames and lexical diversity in the slot position of frames indicating that basically only a small number of frames are potentially useful. We first discuss issues concerning the cross-linguistic applicability of the distributional pattern before turning to the potential usefulness of frequent frames in German acquisition.

Table 4

Mean (*SD*) category accuracy for Standard and Expanded Labeling (token and type) including mean (*SD*) accuracy for random categories: A_x_B frames

Categories	Token Accuracy		Type Accuracy	
	Analysis	Random	Analysis	Random
Standard Labeling	.10 ^a (.06)	.03 ^a (.000)	.11 ^c (.07)	.03 ^c (.000)
Expanded Labeling	.11 ^b (.07)	.03 ^b (.000)	.07 ^d (.06)	.03 ^d (.000)

Note: ^{a,b,c,d}Scores differ significantly (Fisher's Omnibus test, $p < .001$).

Table 5

Mean frame accuracy (*SD*) for Standard and Expanded Labeling (token and type) including mean accuracy (*SD*) of random categories: A_x frames (*N* = 45)

A_x	Token Accuracy		Type Accuracy	
	Analysis	Random	Analysis	Random
Standard Labeling	.39 ^{a,e,g} (.18)	.21 ^a (.00)	.29 ^{b,f,g} (.14)	.21 ^b (.00)
Expanded Labeling	.34 ^{c,e,h} (.18)	.15 ^c (.00)	.25 ^{d,f,h} (.14)	.17 ^d (.00)

Note: ^{a,b,c,d}Scores differ significantly (Fisher’s Omnibus test, *p* < .001).

^eMeans differ significantly (paired t tests, *p* < .001).

^{f,g,h}Means differ significantly (Wilcoxon Signed Ranks test, *p* < .001).

Table 6

Mean token accuracy (*SD*) for three groups of A_x frames

A_x Group	<i>n</i>	Standard Labeling	
		Token Accuracy	Shannon–Weaver
Accurate and diverse ^a	5	.68 (.11)	.85 (.11)
Accurate but lexically specific ^a	4	.71 (.08)	.39 (.18)
Inaccurate and diverse ^a	36	.31 (.10)	.79 (.07)

Note: ^a“Accurate” corresponds to scores equal to or greater than .57. “Inaccurate” corresponds to scores less than .57. “Diverse” corresponds to scores that fall within or above the 1 *SD* range around the mean. “Lexically specific” corresponds to scores that fall below 1 *SD* below the mean.

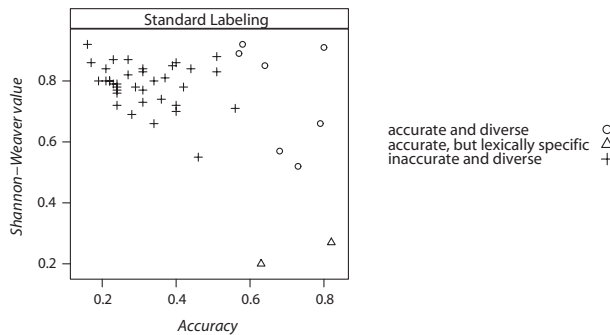


Fig. 5. Diversity (Shannon–Weaver values) as a function of accuracy for A_x frames (Standard Labeling).

4.1. Cross-linguistic applicability

Why did the frequent frames in German child-directed speech gather such inconsistent sets of grammatical categories? Categories that often ended up in the same frame were *noun*, *pronoun*, *main verb*, *auxiliary verb*, and *adverb*. There are at least three explanations. First, this is due to the fact that virtually all definite articles that were registered as a left framing element were used both as a determiner and a pronoun. For instance, a frame like

das_x_nicht categorized verbs as well as adverbs, nouns, and pronouns (e.g., *das_ist_nicht* (verb; *is*); *das_noch_nicht* (adverb; *still*); *das_huhn_nicht* (noun; *chicken*); *das_beide_nicht* (pronoun; *both*)). *Das* was used pronominally if followed by a verb or an adverb, but was used as a determiner if followed by a noun or a pronoun. Second, frames sometimes crossed the boundaries of intonation units within a sentence. Although frames were not allowed to cross utterance boundaries considering the whole sentence as one intonation unit, there were no restrictions for frames within one sentence; that is, speech pauses were ignored as boundaries of intonation units. For instance, the frame *das_x_eine* collected verbs most of the time but gathered adverbs as well. In the latter case the frame crossed a pause: “*Was ist das_da,_eine?*” (“What’s that, a?”). Third, for some frames, low accuracy scores together with high diversity scores might mirror the variability in the relative ordering of the subject and the verb in German. For instance, the frame *ich_x_nicht* categorized, among others, verbs (e.g., *wissen* “to know”) and adverbs (e.g., *wirklich* “really”). Both could occur at either side of the subject of the sentence, which is the right-framing element of the frame. Whenever there is an adverb or a pronoun in the slot position of this frame, the subject *ich* (I) must be preceded by a verb. In contrast, whenever there is a verb in the slot position, the subject *ich* (I) can be preceded by an adverb. The following sentences, which are basically the same speech act for the same event and contain the same lexical items, are possible: “*wirklich, ich_weiß_nicht was du meinst*” (Really, I don’t know what you mean), “*Was du meinst weiß ich_wirklich_nicht*” (What you mean know I really not). Whereas the last two explanations might apply to Dutch frequent frames as well, the first reflects a characteristic of German only.

Furthermore, Mintz (2003) hypothesized that in languages with a more flexible word order co-occurrence patterns at a different level of granularity, for example, at the level of sublexical morphemes, might be more informative. Following this, Erkelens (2009) found evidence for categorization from frequent morpheme frames for 16-month-old children learning Dutch. Evidence for productive morpheme frames in German comes from a corpus study showing young children’s sensitivity to the phonological patterns in word structure (word endings) and their co-occurrence with gender marked articles (Szagun, Stumper, Sondag, & Franik, 2007). Further, frequent morpheme frames in German seem to show higher accuracy scores as compared to frequent word frames (Höhle, personal communication).

4.2. Frequent frames in acquisition

In the case of those frames that exhibit high accuracy but low diversity not only the framing elements remain constant but also the items in the slot position stay rather unchanged. Therefore, these frames are likely to become associated with only one particular lexical item, making them less accessible for use with new items (Bybee, 2007). Evidence that children are less likely to extract lexical frames from the input where they encounter little diversity in the word forms found in that frame has recently been reported by Matthews and Bannard (2010). An indication for the frames occurring highly frequently with one specific word in the slot position derives from the fact that some of the words in the slot position were phonetically shortened (e.g., *denn* (particle, no translation) to ‘*nn*, *ist* (is) to *is* or ‘*s*)

(Bybee, 2001; Jurafsky, Bell, Gregory, & Raymond, 2001). Further evidence for the frames being a fully lexically specific three-word pattern comes from a study by Stoll, Abbot-Smith, and Lieven (2009). They investigated sentence-initial, lexically based patterns in six corpora of German child-directed speech. There is considerable overlap such that five of our frame-based patterns frequently occurred in their corpora, too. Thus, being rather lexically specific formula, these accurate but invariable frames are not going to help with categorization of words.

By contrast, a frame showing both high accuracy and high diversity may be a guide to learn the category of the intervening word. The more types are used within a certain position in a pattern, the more likely it is that a general category will be formed over the lexical items that occur in that position (Bybee, 2007; Onnis, Monaghan, Christiansen, & Chater, 2004). Further, the type frequency of a pattern determines whether it will be used productively and be applied to new forms (Bybee, 1995; Goldberg, 1995; MacWhinney, 1978). As in Mintz (2003) the slots of some of the frames contained simple transitive verbs (*was_x_du*, *was_x_der*, *was_x_die*, *ich_x_das*). But the right-framing element was often only the first part of the verb's argument as the definite articles were used both as a determiner (then followed by a noun or an adjective plus noun) and a pronoun. The frame structure is thus too narrow to capture the whole transitive construction.

Finally, those frames that show low accuracy together with high diversity are rather noisy and misleading constructions. Depending on whether main verbs and auxiliary verbs were grouped (Standard Labeling) or not (Expanded Labeling), accuracy scores differed considerably. Thus, accuracy of frames is sensitive to the level of granularity. Frames are more capable of roughly categorizing verbs as in Standard Labeling than of depicting subtle syntactic categories as in Expanded Labeling.

It seems, then, that accurate and diverse frames might help the child in learning the category of the intervening word, particularly of verbs as most of the frames in the corpus collected verbs. But do children actually use such frames? Establishing the psychological reality of frequent frames in German will require experimental investigations (Erkelens, 2009; Mintz, 2006). Important evidence about children's use of frames might also come from studying their production. The process of discovering categories is complementary to the discovery of productive linguistic patterns (see Matthews & Bannard, 2010), and we might expect that children's use of frames in detecting categories would go hand in hand with their use of those lexical frames in production. Finally, cognitively plausible computational models that utilize co-occurrence statistics (Freudenthal et al., 2009; Monaghan & Christiansen, 2004) will be valuable in exploring how children might actually use the information.

In general what we have discovered with this analysis is that the child utilizing frequent frames in German is faced with a tension between the need for coverage and the need for accuracy—the child will only be able to accurately infer categories for a small part of the vocabulary. Achieving anywhere near full coverage would require them to rely on messier data and risk making a substantial number of categorization errors. The data did not reveal any clear strategy that the child might use to overcome this challenge. It is important to note, however, that distributional information is not the only

information available and the use of other information may help them to distinguish categories from noise. It seems likely, for example, that children pick up those recurring patterns that help in achieving communicative functions. Thus, future research on children's skills of linguistic categorization should focus on communicative function as an essential element which is not considered by distributional analyses that simply rely on sequentially co-occurring items.

4.3 Conclusion

A clear limitation of the present study is that only one child-directed speech corpus contributed to the analysis. Although being relatively dense, the corpus is still only a snapshot of the linguistic environment to which the child is exposed. Nevertheless, the current study on German adds to the body of research showing that no single cue—in our case information about co-occurrences at the word level—provides all necessary information about word class membership (Monaghan, Christiansen, & Chater, 2007; Morgan, Shi, & Allopenna, 1996). Therefore, the child must have to probabilistically exploit a variety of sources that are available in the input.

Notes

1. See the Appendix for a translation of all frames.
2. An analysis of accuracy for x_B partial frames resulted in comparable but again smaller scores ranging from .23 to .36.

Acknowledgments

We would like to thank Frank Binder and Patrick Jähnichen for preparing the corpus, Roger Mundry for statistical guidance, and two anonymous reviewers who provided helpful comments on an earlier version of this paper.

References

- Behrens, H. (2000). *Anleitung zur Kodierung und Disambiguierung von Erwerbsdaten des Deutschen*. Unpublished manuscript. Germany: Max Planck Institute for Evolutionary Anthropology Leipzig.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57, 289–300.
- Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10, 425–455.
- Bybee, J. L. (2001). *Phonology and language use*. Cambridge, England: Cambridge University Press.
- Bybee, J. L. (2007). *Frequency of use and the organization of language*. Oxford, England: Oxford University Press.

- Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using 'frequent frames': What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science, 12*, 396–406.
- Erkelen, M. A. (2009). *Learning to categorize verbs and nouns: Studies on Dutch*. Utrecht, The Netherlands: LOT.
- Freudenthal, D., Pine, J. M., & Gobet, F. (2009). On the utility of conjoint and compositional frames and utterance boundaries as predictors of word categories. *Proceedings of the 30th Annual Conference of the Cognitive Science Society, USA, 08*, 1947–1952.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University Press.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.
- Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*, 183–206.
- Haccou, P., & Meelis, E. (1992). *Statistical analysis of behavioural data*. Oxford, England: Oxford University Press.
- Jurafsky, D., Bell, A., Gregory, M. L., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Benjamins.
- MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development, 174*, 43.
- MacWhinney, B. (2000). *The CHILDES-project: Tools for analyzing talk*. Mahwah, NJ: Erlbaum.
- Matthews, D., & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive Science, 34*(3), 465–488.
- Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91–117.
- Mintz, T. (2006). Finding the verbs: Distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 31–63). Oxford, England: Oxford University Press.
- Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26*, 393–424.
- Monaghan, P., & Christiansen, M. H. (2004). What distributional information is useful and usable in language acquisition? *Proceedings of the 26th Annual Conference of the Cognitive Science Society, USA, 04*, 963–968.
- Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*, 259–305.
- Morgan, J. L., & Demuth, K. (1996). Signal to syntax: An overview. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 1–22). Mahwah, NJ: Lawrence Erlbaum Associates.
- Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Towards a broader conceptualization of bootstrapping. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 263–283). Mahwah, NJ: Lawrence Erlbaum Associates.
- Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing nonadjacent dependencies. *Proceedings of the 26th Annual Conference of the Cognitive Science Society, USA, 04*, 1047–1052.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*, 425–469.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

- Stoll, S., Abbot-Smith, K., & Lieven, E. (2009). Lexically restricted utterances in Russian, German and English child-directed speech. *Cognitive Science*, 33, 75–103.
- Szagan, G., Stumper, B., Sondag, N., & Franik, M. (2007). The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language*, 34, 445–471.
- Zar, J. H. (1999). *Biostatistical analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Appendix

All 45 frequent frames with translation of the framing elements

Frame	Translation
<i>die_x_ist</i>	<i>theFEM(det)_x_is</i>
<i>der_x_ist</i>	<i>theMASC(det)_x_is</i>
<i>Leo_x_mal</i>	<i>Leo_x_particle(no translation)</i>
<i>das_x_der</i>	<i>theNEUT(det/pro)_x_theMASC(det/pro)</i>
<i>was_x_denn</i>	<i>what_x_particle(no translation)</i>
<i>wo_x_die</i>	<i>where_x_theFEM(det/pro)</i>
<i>haben_x_denn</i>	<i>have(3rdpl)_x_particle(no translation)</i>
<i>wo_x_denn</i>	<i>where_x_particle(no translation)</i>
<i>was_x'_n</i>	<i>what_x_particle(no translation)/theMASC(detACC)</i>
<i>wo_x_der</i>	<i>where_x_theMASC(det)</i>
<i>das_x_die</i>	<i>theFEM(det)_x_theNEUT(det/pro)</i>
<i>was_x_das</i>	<i>what_x_theNEUT(det/pro)</i>
<i>das_x_ein</i>	<i>theNEUT(det)_x_aMASC,NEUT(det/pro)</i>
<i>ist_x_das</i>	<i>is_x_theNEUT(det/pro)</i>
<i>guck_x_hier</i>	<i>look_x_here</i>
<i>und_x_macht</i>	<i>and_x_does</i>
<i>wir_x_mal</i>	<i>we_x_then</i>
<i>Leo_x_du</i>	<i>Leo_x_you</i>
<i>wir_x_noch</i>	<i>we_x_still</i>
<i>dann_x_wir</i>	<i>then_x_we</i>
<i>jetzt_x_der</i>	<i>now_x_theMASC(det/pro)</i>
<i>da_x_die</i>	<i>there_x_theFEM(det/pro)</i>
<i>ich_x_das</i>	<i>I_x_theNEUT(det/pro)</i>
<i>jetzt_x_du</i>	<i>now_x_you</i>
<i>oh_x_ist</i>	<i>oh_x_is</i>
<i>das_x_nicht</i>	<i>theNEUT(det/pro)_x_not</i>
<i>ich_x_nicht</i>	<i>I_x_not</i>
<i>ich_x_mal</i>	<i>I_x_particle(no translation)</i>
<i>und_x_haben</i>	<i>and_x_have(3rdPl)</i>
<i>jetzt_x_wir</i>	<i>now_x_we</i>
<i>Leo_x_ist</i>	<i>Leo_x_is</i>
<i>was_x_die</i>	<i>what_x_theFEM(det/pro)</i>
<i>was_x_du</i>	<i>what_x_you</i>
<i>hm_x_ist</i>	<i>hm_x_is</i>
<i>da_x_wir</i>	<i>there_x_we</i>
<i>ist_x_der</i>	<i>is_x_theMASC(det/pro)</i>

Appendix (Continued)

Frame	Translation
<i>was_x_der</i>	<i>what_x_theMASC(det/pro)</i>
<i>da_x_der</i>	<i>there_x_theMASC(det/pro)</i>
<i>was_x_wir</i>	<i>what_x_we</i>
<i>ja_x_ist</i>	<i>yes_x_is</i>
<i>und_x_ist</i>	<i>and_x_is</i>
<i>ist_x_ein</i>	<i>is_x_aMASC,NEUT(det/pro)</i>
<i>ist_x_da</i>	<i>is_x_there</i>
<i>das_x_eine</i>	<i>theNEUT(det/pro)_x_aFEM(det/pro)</i>
<i>das_x_ja</i>	<i>theNEUT(det/pro)_x_particle(no translation)</i>