Cognition 175 (2018) 131–140

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

A universal cue for grammatical categories in the input to children: Frequent frames

Steven Moran[a,*], Damián E. Blasi[a,b], Robert Schikowski[a], Aylin C. Küntay[c], Barbara Pfeiler[d], Shanley Allen[e], Sabine Stoll[a]

[a] Department of Comparative Linguistics, University of Zurich, Switzerland
[b] Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Germany
[c] Department of Psychology, Koç University, Turkey
[d] Universidad Nacional Autónoma de México, Mexico
[e] Department of Cognitive Science and Psychology, University of Kaiserslautern, Germany

ARTICLE INFO

Keywords:
Statistical learning
Nonadjacent dependency
Cross-linguistic language acquisition
Frequent frames
Input patterns
Child-directed speech

ABSTRACT

How does a child map words to grammatical categories when words are not overtly marked either lexically or prosodically? Recent language acquisition theories have proposed that distributional information encoded in sequences of words or morphemes might play a central role in forming grammatical classes. To test this proposal, we analyze child-directed speech from seven typologically diverse languages to simulate maximum variation in the structures of the world's languages. We ask whether the input to children contains cues for assigning syntactic categories in frequent frames, which are frequently occurring nonadjacent sequences of words or morphemes. In accord with aggregated results from previous studies on individual languages, we find that frequent word frames do not provide a robust distributional pattern for accurately predicting grammatical categories. However, our results show that frames are extremely accurate cues cross-linguistically at the morpheme level. We theorize that the nonadjacent dependency pattern captured by frequent frames is a universal anchor point for learners on the morphological level to detect and categorize grammatical categories. Whether frames also play a role on higher linguistic levels such as words is determined by grammatical features of the individual language.

1. Introduction

Humans learn language through exposure to surrounding speech. Speech is rich with distributional regularities encoded in adjacent and nonadjacent sequences, which reflect grammar constraints. Experimental studies suggest that infants are sensitive to dependencies between sequences and they can use general mechanisms of statistical learning to process and acquire language (for a review see Sandoval & Gómez, 2013). Infants can, for example, segment the speech stream into words given only dependencies between adjacent syllables (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). But to attain linguistic proficiency, children must also learn to generalize the behavior of words into grammatical categories, so that they can be used productively in syntax.

The mechanisms that children use to assign and remember grammatical category membership are not well understood. How does a child learn to map words to classes when words are not overtly marked, cross-linguistically, neither lexically nor prosodically? Language-specific phonological cues such as stress or segment length (Cassidy & Kelly, 2001) have been shown to facilitate word category assignment (Monaghan, Christiansen, & Chater, 2007). However, not all languages have phonological cues that accurately predict grammatical categories. So how are they learned? One other promising candidate are structural cues such as neighboring words or discontinuous dependencies, that are indicative of grammatical category.

Words belonging to the same category typically behave similarly in similar morphological and syntactic contexts (Bloomfield, 1933; Harris, 1951). Members of the same class, such as 'noun' or 'verb', can be substituted for one another without changing the grammaticality of an utterance. Presumably, these distributional patterns provide input regarding grammatical function to the learner. Maratsos and Chalkley (1980) propose that adjacent sequences in word cooccurrence distributions are a cue for word categorization. Cartwright and Brent (1997) and Redington, Crater, and Finch (1998) use bigram frequencies from natural language and computer simulations to demonstrate categorical learning effects. And Mintz, Newport, and Bever (2002) show that distributional structures in adjacent dependencies (bigram cooccurrences) successfully categorize nouns and verbs in child-directed speech in four English corpora.

* Corresponding author.
E-mail address: steven.moran@uzh.ch (S. Moran).

https://doi.org/10.1016/j.cognition.2018.02.005
Received 17 February 2017; Received in revised form 18 January 2018; Accepted 6 February 2018
Available online 16 March 2018
0010-0277/ © 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

Table 1
Results from previous studies.

Language (corpus)	Utterances	Mean accuracy		Mean completeness	
		Words	Morphemes	Words	Morphemes
English (Mintz, 2003)	103,191	0.91		0.12	
Chinese (Xiao et al., 2006)	22,137	0.70			
Dutch (Erkelens, 2009)	49,635	0.71			
French (Chemla et al., 2009)	2006	1.0		0.33	
Spanish (Weisleder & Waxman, 2010)	37,588	0.75			
Turkish (Wang et al., 2011)	37,765	0.47	0.91	0.10	0.06
German (Wang et al., 2011)	5685	0.86	0.88	0.07	0.05
German (Stumper et al., 2011)	30,601	0.77			

In addition to adjacent dependencies, nonadjacent dependencies in natural language exist and they can encode grammatical structures. An example is morphosyntactic agreement, e.g. he **is** sleep**ing**. There is ample evidence that infants make use of nonadjacent dependencies in categorizing elements presented between two repetitive surrounding elements (Gómez, 2002; Gómez & Maye, 2005; Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004; Mintz, 2006; Nazzi, Floccia, Moquet, & Butler, 2009; Nazzi, Barrière, Goyet, Kresh, & Legendre, 2011; Onnis, Monaghan, Christiansen, & Chater, 2004; Santelmann & Jusczyk, 1998; Van Kampen et al., 2008). Artificial language learning experiments also show that learners are sensitive to nonadjacent dependencies (Wang & Mintz, 2016). The simplest nonadjacent dependency is the so-called *frame*, a sequence of three elements, like A_B_C, in which A and C predict information about B. In our example of morphosyntactic agreement, only verbs can appear between the auxiliary verb *is* and the progressive suffix *-ing*. Therefore this nonadjacent dependency, or frame, signals the grammatical class of the intervening element.

Mintz (2003: 91) defines the frame as, "two jointly occurring words with one word intervening", and shows that words A and C in frequently occurring frames accurately categorize the grammatical category of word B in English. Across longitudinal corpora of child-directed speech in parent-child dyads, the results are robust as evaluated by measures of accuracy and completeness. In technical terms, accuracy is equivalent to *precision* in Information Retrieval, i.e. true positives/true positives + false positives (aka a Type I statistical error). And completeness is analogous to *recall*, i.e. true positives/true positives + false negatives (aka a Type II statistical error). In plain speak, accuracy measures how precise is the set of elements selected from a sample. For example, you want to select apples from a bag of apples and pears, but you cannot see in the bag. Out of the pieces of fruit you pick from the bag, how many are apples? Completeness measures how many apples you selected from all the apples present in the bag.

Since Mintz (2003), studies of frequent word frames in languages other than English have had mixed results, summarized in Table 1. French and Spanish frames are a robust cue for word categorization, especially for nouns and verbs (Chemla, Mintz, Bernal, & Christophe, 2009; Weisleder & Waxman, 2010). Frames in Dutch, German and Turkish, however, are not accurate (Erkelens, 2008; Stumper, Bannard, Lieven, & Tomasello, 2011; Wang et al., 2011). Erkelens (2008) found that on all levels of analysis, English frames were more predictive than Dutch frames. Weisleder and Waxman (2010: 1098) conclude in their comparison of Spanish and English frames that "the clarity of the distributional information available in frequent frames varies across languages, and within languages it varies across different distributional environments and grammatical form classes". Additionally, different studies using the same methods on different datasets of the same language obtain different results (see the results for German in Table 1). Wang et al. (2011) study word frames in a small corpus of German and found a high degree of accuracy for frames. Stumper et al. (2011), by contrast analyze a much larger corpus of German child directed speech and they find less robust accuracy for word frames.

Hence, it has become a matter of debate whether the nonadjacent dependency captured by the frame is a universally available pattern to children that might aid in categorization. Most studies analyze frames at the word level, i.e. word1_word3. To account for the differences in morphological and grammatical features in typologically different languages, Wang et al. (2011) propose analyzing frames in languages with richer morphology on the morpheme level. They find both Turkish and German frequent morpheme frames are accurate predictors of the target morpheme's grammatical category (morpheme2). This suggests that the morphological complexity of a language might be relevant for the level of granularity of the units where frames are to be found. Whether this finding translates to other languages, however, is an unresolved issue so far and therefore the focus of this paper.

In this paper, we test whether frequent frames are a universally salient nonadjacent distributional pattern at the word and morpheme levels in child-surrounding speech.[1] In Section 2, we describe our data sample, which includes longitudinal corpora from seven typologically diverse languages. Because the corpora differ in size and the languages differ in their morphological complexity, we operationalize a relative frequency measure to make the data comparable. In Section 3, we evaluate frequent frames in child-surrounding speech in each corpus by accuracy and completeness scores and compare the results to previous findings. These measures, however, do not lend themselves to investigating frequent frames cross-linguistically at the level of specific parts of speech. We therefore propose two novel measurements, called global accuracy and global completeness, and test whether certain parts of speech are more accurately captured in frequent frames across languages. In Section 4, we discuss our results and reasons why frequently occurring nonadjacent dependencies captured by frames are indeed a universally available cue for children.

2. Materials and methods

2.1. The corpora

Linguistic diversity poses many challenges for cognitive science (Evans & Levinson, 2009). In language acquisition studies, it is not practical or even possible to test for statistical patterns across all languages. Instead, we simulate linguistic diversity by examining languages which differ maximally in their grammatical structure. To develop a typologically-diverse set of languages, Stoll and Bickel (2013) applied a fuzzy clustering algorithm used by Rousseeuw and Leonard (1990) that takes as input thousands of languages and their typological feature values (e.g. grammatical case, inflection categories, degree of synthesis, inflectional compactness) as encoded in two broad coverage typological databases: the World Atlas of Linguistic Structures (WALS; Haspelmath, Dryer, Gil, & Comrie, 2008) and AUTOTYP (Bickel et al., 2017). The algorithm outputs five clusters of maximally diverse

[1] Data and source code are available at: https://github.com/acqdiv/frequent-frames.

Table 2
Language sample.

Language	Spoken mainly in	Language family	Speakers
Chintang	Nepal	Sino-Tibetan	6000
Inuktitut	Canada	Eskimo-Aleut	34,000
Japanese	Japan	Japanese	128,000,000
Russian	Russia	Indo-European	166,000,000
Sesotho	South Africa	Bantu	5,600,000
Turkish	Turkey	Altaic	70,900,000
Yucatec	Mexico	Mayan	766,000

Table 3
Children in the corpora.

Corpus	Children	Age ranges
Chintang	4	2; 1.9–3; 5.25, 2; 0.29–3; 5.13, 3; 0.14–4; 4.25, 2; 11.2–4; 3.14
Inuktitut	4	2; 6.6–3; 3.2, 2; 0.11–2; 9.5, 2; 6.2–3; 2.26, 2; 9.16–3; 6.12
Japanese	4	2; 11.27–5; 1.23, 2; 11.28–5; 0.17 (×2), 3; 0.1–5; 0.27
Russian	5	1; 3.26–4; 11.0, 1; 4.22–5; 6.26, 1; 6.10–5; 4.18, 1; 11.28–4; 3.14, 3; 1.8–6; 8.12
Sesotho	4	2; 1–3; 0, 2; 1–3; 2, 2; 4–3; 3, 3; 8–4; 7
Turkish	8	1; 0.2–3; 0.3, 0; 7.28–3; 0.24, 0; 8.6–3; 0.14, 0; 8.1–1; 9.28,0; 8.0–2; 4.20, 0; 8.2–3; 0.14, 0; 8.30–3; 0.20, 0; 9.27–2; 9.13
Yucatec	3	1; 11.9–3; 5.4, 2; 0.1–3; 0.29, 2; 1.5–3; 3.11

languages. For each cluster two languages were chosen where longitudinal studies were available (for this study we were forced to use a subset of these languages, see below). Each study contains spontaneous dialogues around a target child. The sessions are captured in recordings that were then transcribed and annotated. The resulting corpora were compiled into the unified ACQDIV database (Moran, Schikowski, Pajović, Hysi, & Stoll, 2016), which can be used to mine for statistical patterns in child surrounding speech.

For the present study, we selected seven languages from the ACQDIV database, which adhered to our experiments' constraints. For example, the Indonesian corpus (Gil & Tadmor, 2007) is not part-of-speech tagged, so we could not use it in any analysis. The Russian corpus (Stoll & Meyer, unpublished) is not morphologically segmented, so we could only use it in our analysis of frequent frames at the word level. Listed in Table 2, the languages in our sample are culturally, geographically, genealogically, and demographically diverse.

Table 3 lists the number of children and their ages in each corpus in our sample. Fig. 1 in the Supplementary Materials in Section 5 shows the age spans for each child. Each corpus's size and the amount of data we analyzed in this study is listed in Table 4. The corpora differ in size and morphological productivity, as shown in their number of utterances, word tokens and morpheme tokens.

To establish that the nonadjacent dependency structure captured by frames is a language universal pattern in child-surrounding speech, we test for predictable categorization in this sample of grammatically diverse languages. The typological diversity that the sample captures is aimed at freeing analyses of input patterns from bias from particular grammatical structures. For example, some grammatical phenomena have been claimed to predispose languages towards frames. On the word level, English frequent frames include "you_it", "the_one", and "put_in" (Mintz, 2003). The first frame is only possible in languages that do not allow pro-drop; the second presupposes that the language has articles; and the third that it contains prepositions. Certain word orders or stricter word order is also potentially predictive because it determines if nouns can easily target verbs, e.g. (NP V NP). The languages that we use in our analysis capture linguistic diversity – e.g. different word orders, pre-or postpositions, languages with and without pro-drop, languages with and without articles – hence statistical regularities identified in this sample of languages are more likely to be universal.

Table 5
Typological features.

	Synthesis			
Language	Word order	Noun	Verb	Adposition
Chintang	SOV	Mid	High	None
Inuktitut	SOV	High	High	None
Japanese	SOV	Low	Low	Post
Russian	SVO	Mid	Mid	Prep
Sesotho	SVO	Mid	High	None
Turkish	SOV	Mid	High	Post
Yucatec	VOS	Mid	High	Prep

As summarized in Table 5, the languages in our sample exhibit morphology with varying degrees of synthesis and fusion. Morphology may also play a role in frame creation. For example, a language with both prefixes and suffixes might capture (A- _ -C) frames. There may also be agreement in the right position of a frame, (head-dependent _ -AGR), e.g. (doggy(Masculine) ADJ -Masculine). The degree of morphological synthesis may also play a role. For example, the more synthesis, the higher the probability for affixes targeting affixes, as in Chintang (-u'3P' _ -a'IMP').

2.2. Extraction procedure

Our frame identification procedure extracts utterances of child-directed and child-surrounding speech from each corpus. Since we do not have annotations for directedness in all corpora, child-directed utterances are approximated as utterances made by all adult speakers. Each utterance is split into sequences of words or morphemes and their part-of-speech labels. From these n-gram sequences, all trigrams are extracted. Each trigram, say A_B_C, is a frame and a potential frequent frame. A *frame* is an ordered pair of items, in our study space-delimited words and expert-annotated morphemes, with a corresponding element intervening. A *frame-based category* is the set of frames that contain the

Table 4
Corpus and analysis size.

Language (corpus)	Utterances	Words		Morphemes	
		Total	Analyzed	Total	Analyzed
Chintang (Stoll et al., unpublished)	396,412	987,120	473,918	1,594,829	814,076
Inuktitut (Allen, unpublished)	46,680	73,255	23,164	37,781	8673
Japanese (Miyata, 2012)[a]	271,868	821,106	514,344	666,748	376,934
Russian (Stoll & Meyer, unpublished)	828,041	2,033,755	1,316,234	NA	NA
Sesotho (Demuth, 2015)	69,530	237,112	83,514	329,347	112,630
Turkish (Küntay et al., unpublished)	400,836	1,136,332	938,955	300,907	272,459
Yucatec (Pfeiler, unpublished)	91,825	257,496	89,219	198,761	84,928

[a] Based on Miyata & Nisisawa (2009, 2010) and Nisisawa & Miyata (2009, 2010).

same framing elements (Mintz, 2003), i.e. all occurrences of, say, A_C. The intervening element, the so-called *target*, is element B and its grammatical label.

Frame-based categories do not by definition include utterance boundaries as framing elements (#_B and B_# in the sequence #ABC#) nor do frames cross utterances, e.g. C_E in utterances ABC and DEF (Mintz, 2003: 96). In this study, we focus on lexical frequent frames' unique contribution in the input to children and we do not take utterance boundaries as framing elements. Support comes from an artificial language learning experiment by Wang and Mintz (2016), who show that pauses and edges are not needed to learn nonadjacent dependencies. We do note, however, that position salience should be investigated cross-linguistically; see Section 5.6 in the Supplementary Materials. Consider the study by Freudenthal, Pine, Jones, and Gobet (2013), who investigate why children form a productive noun category earlier than a verb category. Their results suggest that utterance final position is a more accurate predictor of nouns than frame-based categories that include utterance boundaries.

Each frame-based category in our analysis is assigned a modal grammatical category following Weisleder and Waxman (2010). That is, the grammatical category of targets in a frame-based category are tallied and the grammatical category with the most occurrences is assigned as the modal category. For example, the Chintang word frame-based category, aŋ_lo, occurs 96 times with verb as target and once with a noun and once with an adverb. Therefore the modal category assigned is verb, with 98% of occurrences. When we average across modal categories at the word level, the distribution ranges from 55% in Russian to 99% in Inuktitut with a cross-corpus mean of 81%. At the morpheme level, the distribution is over 94% for all corpora with a cross-corpus mean of 97%.

In the development of the ACQDIV database of corpora, each corpus was annotated for parts of speech independently at the word and morpheme levels. This procedure was undertaken by mapping richly annotated language-specific phenomena to a normalized set of parts of speech. For example, Sesotho has a rich noun class system with over a dozen glosses annotated by "nˆ" followed by a sequence of digits, e.g. nˆ3, nˆ5, nˆ10, to indicate specific noun classes (Demuth, 2015). We map this Sesotho-specific labelling system to the label *Noun*, thus making the richly annotated pluralization strategies of nouns in Sesotho comparable with nouns in other languages in our sample. Our normalization of parts of speech results in a set of twelve categories that we use in our analyses: adjective, adverb, auxiliary, conjunction, interjection, noun, numeral, postposition, preposition, pronominal-demonstrative, particle and verb. Additionally, dependent morphemes have the label prefix or suffix. Labels on the word and morpheme levels apply to stems and not word forms.

Our approach differs from previous studies that use sets of grammatical category labels, either "standard" or "extended", which range in number and type. For example, standard labelling by Mintz (2003: 97) includes ten parts of speech and extended labelling contains a few additional distinctions – nouns are divided into nouns and pronouns, and verbs into verbs, auxiliaries and copula – resulting in fourteen categories total. The corpora in this study also include fourteen labels (see Table 14 in the Supplementary Materials). We do not analyze a standard versus extended labelling scheme because a larger number of category distinctions should, if anything, decrease the overall accuracy of frames because the labelling of targets is more precise, so there is a greater distribution of types. In other words, fewer label types should increase the accuracy of categorization. In fact, Stumper et al. (2011: 1194) find a statistically significant difference between standard and expanded labelling schemes for accuracy scores. Hence, we decided to err on the side of caution and go with a larger set of finer-grained distinctions because if a pattern is found there, then it is more likely to reflect an actual signal. Consider the most extreme example of labelling all word forms with the same label; categorization accuracy would always be perfect. Therefore we think it pertinent to use a relatively large, transparent and cross-linguistically applicable grammatical labelling scheme in our analyses.

Lastly, during our data extraction procedure we calculate the categorization accuracy of each bigram, i.e. the conditional probability of any two adjacent words or morphemes. We exclude from our analyses any frame whose A or C element is a better predictor of B's grammatical category in the nonadjacent dependency A_B_C. This is because the investigation in this study is about the potential of nonadjacent dependencies as distributional patterns that accurately categorize grammatical categories. If an adjacent dependency within a frame is a more accessible or reliable cue than the frame itself, then we do not include the frame in our analyses. For example, consider a hypothetical corpus in which most occurrences of the word 'nice' is followed by a noun. If the accuracy of the bigram ('nice', N) is greater than the accuracy of a frame-based category that contains nice_x, then we assume the bigram is a better classifier of the part-of-speech than the frame and do not include the frame in our analyses. Our results are in line with Chemla et al. (2009), who show that derived categories from frequent frames are more accurate than those of bigram cooccurrences, e.g. A_B_x or x_A_B. Lastly, our decision to exclude these frames may also be cognitively motivated. Experimental evidence shows that the unreliability of adjacent cues spurs children to consider nonadjacent cues (Gómez, 2002).

2.3. Operationalization of frequency

Determining the frequency of frequent frames has been approached in a few different ways, but the constraints of these studies are similar: input data are child-directed speech extracted from parent–child dyads. In the seminal frequent frames study by Mintz (2003), six English corpora from the CHILDES database (MacWhinney, 2000) were used as input (average size 17,199 utterances). For each corpus, the 45-most frequently occurring frames were selected as frequent frames, which satisfied two criteria in the experiment: that the frames were frequent enough to be noticeable and that the frequent frames should include a variety of intervening target words, which could be categorized together (Mintz, 2003: 96). Although their corpora differ in size, studies by Erkelens (2008, 2009), Weisleder and Waxman (2010), Wang et al. (2011), and Stumper et al. (2011) use the 45-most frequent frames per corpus.[2]

Xiao, Cai, and Lee (2006) analyze two small corpora (9403 and 12,734 utterances) and operationalize a frequent frame as any frame occurring at least 15 times. By contrast, instead of fixing the number of frames a priori, Chemla et al. (2009) first evaluate all frames in their corpus (2006 utterances) and established a frequency threshold by evaluating performance iteratively on a successively larger number of frame-based categories. They then relaxed the inclusion criteria to select a set of frames that exclusively contained words from only one category, thereby achieving categorization accuracy of 100% and identifying frame-based categories with a single category.

The ACQDIV corpora create two new challenges for operationalizing frequency in frame evaluation. First, the seven corpora in this study range by orders of magnitude in size and predefining the threshold of what is frequent penalizes cross-linguistic comparison and may introduce bias into the evaluation metrics. Second, the morphologies of languages in our sample are intended to differ maximally. Therefore, the definition of what a word is differs greatly. Differences in morphosyntax mean that each language will have more or less adjacent and nonadjacent dependencies, depending on how the speech stream is segmented (see Section 4).

[2] Weisleder and Waxman (2010) observe that frames that occurred at least 45 times in their corpora surpassed 5% of the proportion of the total number of frames per corpus. Results were robust and comparable when their analysis was set to the top 25-most frequent frames.

Because our corpora differ in size and the languages differ in morphosyntactic productivity, we introduce an operationalization of frequency so that we can control for corpus size. We do so by calculating the frequency of each frame-based category within each corpus in relative proportion to the size of the corpus (the total number of trigrams). Most frame-based categories, both at the word and morpheme levels, occur infrequently. For word frames we found through inspection that the majority of frame-based categories occur with a relative frequency of less than 0.0005%. For morphemes, which are greater in total number of tokens, this relative frequency was 0.001%. We set these values as our variable thresholds and evaluate per corpus the frames above these thresholds. Interestingly, accuracy measures for a fixed threshold for frequency (top-45 frames as per Mintz, 2003) and a variable threshold (as we propose here) are not statistically significantly different, as shown below in Section 3.

2.4. Evaluation

2.4.1. Accuracy

To evaluate how useful frequent frames might be as a bootstrapping cue for the learner, Mintz (2003), and authors of subsequent studies on frequent frames (Chemla et al., 2009; Erkelens, 2009; Stumper et al., 2011; Wang et al., 2011; Weisleder & Waxman, 2010; Xiao et al., 2006), calculate the summary measures *accuracy*, and to a lesser degree, *completeness*, as used first by Cartwright and Brent (1997: 144). Both accuracy and completeness measures range from 0 to 1.

Accuracy measures the categorization success of a frame-based category. In other words, are the linguistic units that occur in the middle of a frame predictive for one specific part-of-speech? Accuracy is defined as the number of *hits* divided by the number of *hits* + *false alarms* (Mintz, 2003: 96). A hit is equivalent to a true positive and a false alarm to a false positive (Type I error). To calculate the accuracy of a frame then, for each frame-based category, all pairs of target words (or morphemes) are compared and if their part-of-speech is the same, a hit is recorded, otherwise it is a false alarm. As such, accuracy is a pairwise measure that is sensitive to the size of the frame-based category. For example, if the frame *I_you* appears three times with intervening words: *love* (verb), *hear* (verb), *not* (negation), then there is one hit (*love* & *hear*, verb = verb) and two false alarms (*hear* & *not*, verb ≠ negation; *love* & *not*, verb ≠ negation), resulting in: 1 (hit)/1 (hit) + 2 (false alarms) or 1/3. The frame's accuracy for categorization is therefore 0.33 (33%). Note that overall accuracy scores as previously reported are an average over the accuracy of each frame-based category, which means that smaller frame-based categories contribute disproportionately to the accuracy score (we return to this issue in the Supplementary Materials in Section 5.5).[3] A perfectly accurate frame-based category has an accuracy score of 1 (100%) and categorizes a single part-of-speech for each and every occurrence of the frame.

2.4.2. Completeness

Completeness compares all target types captured by a frame-based category with all types found in the set of frequent frames (Mintz, 2003: 97). It is defined as the sum of the number of *hits* divided by the number of *hits* + *misses* (Mintz, 2003: 97). A hit is equivalent to a true positive and a miss is analogous to a false negative (Type II error). A completeness score of 1 means that all types are captured by the frame.

In the completeness approach taken by Mintz (2003), parts of speech in each frame-based category are pooled together for evaluation. For example, if a frequent frame contains both verbs and negation, then its verb and negation types are compared against the set of all verbs and negators captured by the 45-most frequent frames in the analysis. In the accuracy example, above, the frame *I_you* appears three times with intervening words: *love* (verb), *hear* (verb), *not* (negation). It therefore captures two verb types (love and hear) and one negator (not). If in the set of frequent frames there are eight verb types (love, hear, think, hug, fly, wash, do, play) and one negation type (not), then the completeness value for the frame capturing (love, hear, not) is 3 (hits)/3 (hits) + 6 (misses), resulting in a completeness score of 3/9 or 0.33. That is, the types in the frame-based category capture 33% of the total number of types under investigation.

The completeness metric used by Mintz (2003) captures how representative the frame-based category is in relation to all types captured by the frequent frames, but not in the whole corpus. In other words, how complete is each frequent frame in categorizing grammatical class types found amongst all types captured by the set of frames determined to be frequent? In order to make our results comparable, we also evaluate completeness in this fashion. It is unclear, however, why the sum of frequent frames should be an informative measure of the input to the child. It seems more intuitive to us to ask how good frequent frames are at categorizing parts of speech in relation to the total input the child receives, instead of just the most frequent frames that s/he hears. Hence, we propose two novel evaluation metrics for categorization, global accuracy and completeness, which take all available input to the child into account.

2.5. Global accuracy and completeness

To calculate global accuracy and completeness, first we determine the modal category of each frame-based category (discussed above). Second, we pool together all target types found in frequent frames by their modal category. Third, we calculate global accuracy and completeness measures, specified below, per grammatical category in regard to all types of that category found in frame-based categories and in utterances in the child-directed speech available in each corpus.

Global accuracy is the mean accuracy across frequent frames of the same modal category. We compute global accuracy by averaging the individual accuracy of each frequent frame that categorizes the same modal category, according to their frequency of occurrence. In other words, if the accuracy of frame i is a_i and f_i its frequency, then global accuracy is computed as $A = \sum_i a_i f_i$ where i ranges over all frequent frames with the same modal category. Thus, we calculate the probability of picking any frequent frame *token* and finding that the slot-filling token has a category corresponding to its modal category (for a given category).

Global completeness takes all of the aggregated frames of a specific category (e.g. all modal verb frames) and tests completeness against that category's types, as found in the whole corpus. Therefore we evaluate how complete are frequent frames at capturing specific grammatical categories in the input to the child represented in our corpora. For instance, say out of a set of 50 frequent frames, 25 have the modal category *noun*. We calculate the mean accuracy of those 25 frame-based categories at predicting *noun* (global accuracy). Next, we pool all noun types that occur in the 25 frames and calculate the proportion of noun types that they capture in regard to all noun types that appear in the child-directed speech (global completeness). Lastly, we plot global accuracy vs global completeness to illustrate how well the nonadjacent dependency encoded by frequent frames categorizes each grammatical class, thus showing how accurately and how many types in the lexicon the child may be able to categorize if s/he uses frequent frames.

3. Results

In Analysis 3.1, we test the accuracy and completeness of frequent frames in input to individual children, both with a fixed threshold of 45 frequent frames and with a variable threshold of frequency determined by our operationalization of relative frequency across corpora. This analysis illustrates how an approach to relative frequency of frame occurrence captures differences in corpus sample size and

[3] We thank an anonymous reviewer for bringing this point to our attention.

Table 6
Frequent word frames in Chintang.

		Accuracy	SD	Completeness	SD	Frames	Min	Max	Median
Fixed threshold	LDCh1	0.66	0.25	0.03	0.02	45	27	1311	40
	LDCh2	0.61	0.27	0.03	0.02	45	29	1434	42
	LDCh3	0.55	0.26	0.03	0.02	45	24	478	36
	LDCh4	0.54	0.21	0.03	0.02	45	36	592	56
Variable threshold	LDCh1	0.64	0.25	0.04	0.02	39	29	1311	41
	LDCh2	0.62	0.28	0.03	0.02	41	32	1434	42
	LDCh3	0.54	0.27	0.04	0.03	33	30	478	39
	LDCh4	0.55	0.23	0.04	0.02	35	42	592	66

Table 7
Frequent word frames in Russian.

		Accuracy	SD	Completeness	SD	Frames	Min	Max	Median
Fixed threshold	Child1	0.49	0.25	0.05	0.03	45	96	852	130
	Child2	0.49	0.24	0.05	0.03	45	132	1050	182
	Child3	0.52	0.23	0.05	0.03	45	65	503	94
	Child4	0.47	0.23	0.04	0.03	45	78	560	132
Variable threshold	Child1	0.50	0.26	0.03	0.02	67	84	852	116
	Child2	0.48	0.24	0.04	0.03	51	127	1050	177
	Child3	0.51	0.22	0.04	0.02	50	61	503	85
	Child4	0.46	0.22	0.05	0.03	44	82	560	137

morphosyntactic idiosyncrasies of each language in our sample. We also test whether there is a significant difference between frequent frames to individual children versus data pooled across children. In Analysis 3.2, we evaluate accuracy and completeness of frequent frames at the word and morpheme levels using aggregated child-surrounding speech per language. The results led us to ask whether all or a subset of grammatical categories are accurately captured by frequent frames. In Analysis 3.3, we test each grammatical category in each corpus for global accuracy and completeness. The Supplementary Materials in Section 5 contain additional data and illustrative plots for each analysis.

3.1. Analysis: individual children versus pooled data

In previous studies, frequent frames are evaluated within language-specific corpora of parent-child dyads (Chemla et al., 2009; Erkelens, 2009; Mintz, 2003; Stumper et al., 2011; Wang et al., 2011; Weisleder & Waxman, 2010; Xiao et al., 2006). These studies investigate the input to individual children. We test whether the results between the two conditions (individual vs pooled) differed in any meaningful way.

The Chintang and Russian corpora (Stoll et al., unpublished; Stoll & Meyer, unpublished) contain the densest data samples in our database and therefore were chosen for detailed analysis of individual children. Each corpus contains four target children. The two samples also capture cultural differences in child-rearing. The Chintang spend the majority of their time outdoors and children are raised in a village setting in which the community is involved in child-rearing. The sample contains numerous individuals and results in rich child-surrounding speech. This is in contrast to a more so-called western style of bringing up children, as is captured in the Russian corpus.

Tables 6 and 7 show fixed and variable operationalizations of frequent word frames in Chintang and Russian. For both languages and across individual children at both frequency thresholds, frame accuracy is between 46% and 66% and completeness is low. We give the standard deviations (SD) and the number of frequent frames per child (Frames) and their minimal (Min), maximum (Max), and average (Median) number of occurrences.

Table 8 shows fixed and variable operationalizations of frequent morpheme frames in Chintang. We do not include Russian because the corpus is not segmented into morphemes. The analysis shows that frequent morpheme frames in Chintang are much more accurate indicators of the intervening target morphemes' syntactic category than in frequent word frames in Chintang. The accuracy of these frequent frames is on average over 90%.

To test whether there is a statistically significant difference between our results at the level of individual children and pooled child-directed speech, we used R (R Development Core Team, 2016) to construct a linear model of accuracy as a function of individual child and child-directed speech. Statistical tests show that pooling child-directed speech has no effect: Chintang words and morphemes, $F(4,220) = 1.798$, $p = 0.13$ and $F(4,220) = 1.965$, $p = 0.10$,

Table 8
Frequent morpheme frames in Chintang.

		Accuracy	SD	Completeness	SD	Frames	Min	Max	Median
Fixed threshold	LDCh1	0.95	0.09	0.08	0.07	45	202	2610	269
	LDCh2	0.93	0.13	0.08	0.06	45	226	3138	362
	LDCh3	0.92	0.15	0.07	0.07	45	202	3159	280
	LDCh4	0.89	0.20	0.09	0.06	45	249	3806	387
Variable threshold	LDCh1	0.94	0.13	0.07	0.07	55	175	2610	261
	LDCh2	0.92	0.15	0.06	0.06	59	188	3138	298
	LDCh3	0.92	0.15	0.07	0.06	49	174	3159	273
	LDCh4	0.88	0.20	0.08	0.06	51	237	3806	370

Table 9
Frequent word frames.

	Accuracy	SD	Completeness	SD	Frames	Min	Max	Median
Chintang	0.57	0.24	0.04	0.02	33	90	2720	118.00
Inuktitut	0.98	0.11	0.03	0.01	37	2	3	2.00
Japanese	0.82	0.21	0.02	0.02	97	67	915	106.00
Russian	0.44	0.22	0.04	0.03	48	234	1485	310.00
Sesotho	0.83	0.23	0.01	0.01	107	8	163	12.00
Turkish	0.62	0.20	0.08	0.08	15	34	318	48.00
Yucatec	0.78	0.28	0.01	0.01	133	3	41	3.00

Table 10
Frequent morpheme frames.

	Accuracy	SD	Completeness	SD	Frames	Min	Max	Median
Chintang	0.95	0.09	0.08	0.07	60	517	7940	779.00
Inuktitut	0.93	0.16	0.02	0.01	100	5	43	6.50
Japanese	0.98	0.04	0.02	0.03	187	83	1943	157.00
Sesotho	0.97	0.12	0.04	0.04	88	66	1358	109.50
Turkish	0.88	0.17	0.01	0.01	835	21	1000	37.00
Yucatec	0.90	0.18	0.01	0.02	153	20	584	34.00

respectively, and Russian words ($F(4, 220) = 1.164$, $p = 0.32$). We also tested a linear mixed effects model with accuracy as a function of individual children with corpus as a random effect (Chintang words $t = -0.591$, Chintang morphemes $t = -1.16$, Russian words $t = 1.885$). Both models show the variable individual children is not significant. Therefore we pool child-surrounding speech across children in our analysis of frequent frames.

3.2. Analysis: frequent frames in seven languages

Table 9 shows the accuracy and completeness scores for frequent word frames in each corpus (relative threshold 0.0005%). The overall accuracy and completeness of word frames is not very high. Accuracy ranges from a low of 0.44 in Russian to a high of 0.98 in Inuktitut. Accuracy in Inuktitut is the outlier, but frequent frames occur very rarely. There is a minimum of two and a maximum of three occurrences of repeated frames in the entire corpus. In this light, the high accuracy is not very impressive. Presumably, frames at the word level are not very useful to the child for learning. Also, due to the orthographic tradition in Sesotho, many orthographic words are considered dependent morphemes under a linguistic analysis, which would increase the accuracy of word level frames in this corpus.

Table 10 shows the accuracy and completeness scores for frequent morpheme frames in each corpus (relative threshold 0.001%). In contrast to word frames, morpheme frames in our data are highly accurate predictors of the syntactic categories of target forms. Accuracy scores are 0.88 and above for all languages. Completeness scores are similar to the word frames, i.e. there is not a substantial increase in recall in relation to higher precision.

3.3. Analysis: global accuracy and global completeness

We have shown that frequent frames at the word level are not universally good predictors of syntactic categories, but at the morpheme level they are. Now we ask how globally accurate and complete frequent frames are at categorizing individual parts of speech. We report the results for nouns and verbs at the word and morpheme levels. For the full analysis of all parts of speech, see the Supplementary Materials.

Our results show that frequent frames at the word level are not cross-linguistically accurate predictors of syntactic category for nouns and verbs, as shown in Table 11. The column *Frames* indicates the sum of the occurrences of frequent frames for a particular syntactic category. Japanese has high accuracy for verbs (0.95) and Sesotho has accuracy scores of 0.89 for nouns and 0.87 for verbs. The other languages have accuracy scores below 0.8. Inuktitut is our smallest corpus and it contains the least number of occurrences of frequent noun and verb frames. Inuktitut has the richest morphological system in our language sample, so the number of frequent frames and their number of occurrences is very low and they have high accuracy.

Table 11
Global accuracy and completeness of words (nouns and verbs).

	Corpus	POS	Accuracy	Completeness	Frames
1	Chintang	N	0.23	0.02	234
2	Inuktitut	N	1.00	0.02	6
3	Japanese	N	0.72	0.16	893
4	Russian	N	0.43	0.05	937
5	Sesotho	N	0.89	0.05	81
6	Turkish	N	0.48	0.02	139
7	Yucatec	N	0.75	0.06	120
8	Chintang	V	0.77	0.05	1447
9	Inuktitut	V	1.00	0.02	13
10	Japanese	V	0.95	0.12	628
11	Russian	V	0.54	0.03	690
12	Sesotho	V	0.87	0.05	62
13	Turkish	V	0.70	0.01	95
14	Yucatec	V	0.70	0.04	96

Our analysis of global accuracy and completeness shows frequent morpheme frames are very accurate predictors of noun and verb stems across the languages in our sample; see Fig. 8 in Section 5.5 in the Supplementary Materials. Noun and verb recall is also high, capturing a large portion of the total types in the corpora. Therefore, these categories may be informative for children in grammatical class assignment. Nouns and verbs may also be accurate categories because they are cross-linguistically more salient semantically, than for instance adjectives or other linguistic waste-bucket categories (cf. Payne, 1997: 63). Cross-linguistically, we might also expect that nouns and verbs correlate more frequently with phonological cues, morphosyntactic position and word class, particularly if these signals are used in language acquisition (see Section 4).

As illustrated in Table 12, at the morpheme level we do find strong support for predictable nonadjacent dependencies in child-directed speech for nouns and verbs. Again, Inuktitut has so few frames that we do not focus on it. All other languages have accuracy scores above 0.86. Noun and verb frames at the morpheme level are highly accurate predictors of grammatical categories.

Table 12
Global accuracy and completeness of morphemes (nouns and verbs).

	Corpus	POS	Accuracy	Completeness	Frames
1	Inuktitut	N	1.00	0.01	3
2	Japanese	N	0.93	0.06	291
3	Sesotho	N	0.97	0.14	105
4	Turkish	N	0.91	0.25	601
5	Yucatec	N	0.86	0.30	312
6	Chintang	V	0.99	0.51	479
7	Inuktitut	V	0.80	0.13	26
8	Japanese	V	0.97	0.44	422
9	Sesotho	V	1.00	0.62	448
10	Turkish	V	0.97	0.60	479
11	Yucatec	V	0.98	0.44	262

4. Discussion

We analyzed accuracy and completeness scores for word and morpheme frames in child-surrounding speech (in individual children and pooled across children) in longitudinal child language acquisition corpora from seven unrelated and typologically diverse languages. In accord with the observation from aggregated results across previous studies that focus on single languages, our analysis shows that frequent word frames are not a universally available nonadjacent dependency pattern in child-surrounding speech (Chemla et al., 2009; Erkelens, 2009; Mintz, 2003; Stumper et al., 2011; Wang et al., 2011; Weisleder & Waxman, 2010; Xiao et al., 2006).

However, we do find cross-linguistically that morpheme frequent frames are very accurate predictors of grammatical categories, particularly for nouns and verbs. Verb frames also have high completeness scores. Completeness ranges from 0.44 (i.e. over one-third of verb types in the corpus) to 0.62, except for Inuktitut (0.13), see discussion below. The privileged position of nouns and verbs would seem to invite children to start exploring the part-of-speech system from these categories or from 'proto-part-of-speech' closely matching them (cf. Bar-Sever & Pearl, 2016). What effect frequent frames have on part-of-speech acquisition in general remains an important desideratum for future research.[4] Our results extend the observation by Wang et al. (2011) – morpheme frames in German and Turkish are more accurate than word frames – and show that cross-linguistically, the nonadjacent dependency encoded in the frame is a universally salient distributional pattern in child-surrounding speech. But why are frames more accurate at the level of morphemes than words? And what makes a frame frequent?

In the literature on frames, language-specific morphosyntactic differences have been proposed as the reason for frequent frames and their accuracy, or lack thereof. Stumper et al. (2011: 1198) suggest word frame-based categories in German, as in English, Dutch and French, have function words as framing elements; thus given the high frequency of closed-class words in these languages, there tend to be more frequent frames.[5] Xiao et al. (2006) observe that a high degree of homophony plays a role in word frame frequency in Chinese. In Spanish, pro-drop reportedly increases the likelihood of repetition in word frames (Weisleder & Waxman, 2010). And relatively free word order seems to play a role in the number of exact repetitions in general, so that frames occur more often in English, which has a stricter word order than Russian (Widmer, 2015). Mintz (2003) suggests the morpheme level may capture statistical regularities better than words in morphologically-rich languages with freer word order.

Table 13
Morpheme and word types and their accuracy.

Language	Word types	Accuracy	Morpheme types	Accuracy
Chintang	51,180	0.57	5518	0.95
Inuktitut	12,140	0.98	734	0.93
Japanese	20,746	0.82	7525	0.98
Sesotho	5517	0.83	2437	0.97
Turkish	61,277	0.62	4034	0.88
Yucatec	16,626	0.78	2612	0.90

Although morphosyntactic phenomena like prepositions, articles, etc., highlight areas of grammar that contain frequent and accurate frames in particular languages, our results suggest a simpler solution: that frequency and accuracy of nonadjacent dependencies boils down to less types and more cohesion. At levels of grammar where there are less types and stricter order on linear sequences, we expect to find more frequent and more accurate adjacent and nonadjacent dependencies. In our data sample, there are fewer morpheme types than word types in each language. Compare the number of morpheme and word types and their corresponding accuracy scores in Table 13. A reduction in the number of types increases accuracy. Ancillary evidence is found in the study by St Clair, Monaghan, and Christiansen (2010: 346), who observe that in their corpora bigrams contain more types than trigrams, and that adjacent bigram dependencies are less accurate predictors of the target form's grammatical category than frames.

Inuktitut is an exception, but we believe this is due to its small corpus size (23,164 words, but only 8673 labeled morphemes). As we pointed out above, word frame accuracy in Inuktitut is extremely accurate (0.98), however, in the 37 frame-based categories identified by our operationalization of frequency, no frame is repeated more than three times, with a median value of two. Given such few frame repetitions at the word level, it is not surprising that accuracy is so high. Inuktitut is a highly polysynthetic language. It has over 400 productive affixes and clitics which attach to verbal, nominal or uninflected particle roots, allowing the part-of-speech to change several times within one word.

Why should a reduction in the number of types reflect constraints on the frequency, accuracy, and completeness of nonadjacent dependencies? First, morphologically-rich languages have more word forms, therefore more word types, so accuracy is lower at the word level. Second, accuracy is better at the morpheme level independently of language type because there is stronger cohesion between morphemes than words (words generally have a less strict linear order than morphemes cross-linguistically). Third, these observations are less useful the more a language tends towards isolation because in the extreme case there is no difference between the morpheme and word levels. For example, word frames are accurate categorization patterns in English (Mintz, 2003) because English is a morphologically analytical language. Additional evidence is provided by Wang et al. (2011): in German, the mean accuracy of words and morphemes are nearly the same, but in Turkish, which is morphologically-richer, accuracy scores are much higher in morpheme frames.

If a reduction in types increases frequency in distributional patterns encoded by adjacent and nonadjacent dependencies (in a corpus of the same size), then we expect that accuracy scores at the level of phonotactics are even higher because languages have less phoneme and syllable types than they do morpheme and word types. Also, cohesion at

[4] We agree with one reviewer that our analysis is limited by the size of our corpora and how the data is distributed across the age ranges of children in our language sample. However, stating that there are differences in the speech directed at children of younger and older ages in our sample makes a cross-linguistic generalization that we have not tested empirically. This is an interesting research question to test in the future, which would add to the debate of whether infant- or child-directed speech shares universal properties and it might also shed light on how child-directed speech changes as children get older.

[5] Stumper et al. (2011: 1197) also investigate how evenly target types are distributed across frames using Shannon-Weaver values as a diversity measure. Values ranged from 0.18 to 0.95 (0–1 scale) and they bin the results into categories: accurate and diverse, accurate but lexically specific, inaccurate and diverse. We leave an investigation into frame target diversity for future research.

the phonological level is greater than in morphology and syntax. One way to test this is to investigate accuracy at the phonotactic level. If phonotactic units, insofar as they represent morphologically relevant phoneme combinations, turn out to be equally good or even better classifiers than conventional frequent frames, this would seriously challenge some of our views on the interaction of different linguistic levels in language acquisition. We leave this question for future research.

Contributions

SM, SST designed the research. SM performed the research. DB contributed new analytical tools. SM, SST, RS, DB analyzed the data. SST, AK, SA, BF provided data. SM, SST wrote the paper.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement n° 615988 (PI Sabine Stoll). Many thanks to Katherine Demuth for providing the Sesotho data and to Gabriella Hermon and Caroline Rowland for fruitful discussions. We also warmly thank Lisa Pearl, Daniel Swingley and an anonymous reviewer for helpful comments and suggestions on an earlier draft of this paper. We dedicate our study to Melanie Widmer.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2018.02.005.

References

Allen, S. (unpublished). Allen Inuktitut Child Language Corpus.
Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321–324.
Bar-Sever, G., & Pearl, L. (2016). Syntactic categories derived from frequent frames benefit early language processing in English and ASL. In J. Scott, & D. Waughtal (Eds.). *Proceedings of the 40th annual Boston University conference on language development (BUCLD)* (pp. 32–46). Somerville, MA: Cascadilla Press.
Bickel, Balthasar, Nichols, Johanna, Zakharko, Taras, Witzlack-Makarevich, Alena, Hildebrandt, Kristine, Rießler, Michael, … Lowe, John B. (2017). The AUTOTYP typological databases. Version 0.1.0 https://github.com/autotyp/autotyp-data/tree/0.1.0.
Bloomfield, L. (1933). *Language.* Chicago: The University of Chicago Press.
Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63*(2), 121–170.
Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin & Review, 8*(3), 519–523.
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*(3), 396–406.
Demuth, K. (2015). Demuth Sesotho corpus < http://childes.talkbank.org/access/Other/Sesotho/Demuth.html > .
Erkelens, M. A. (2008). Restrictions of frequent frames as cues to categories: The case of Dutch. In H. Chan, E. Kapia & H. Jacob (Eds.), *A supplement to the proceedings of the 32nd Boston University conference on language development (BUCLD)* (pp. 1–12). Boston, USA.
Erkelens, M. A. (2009). Learning to categorize verbs and nouns: Studies on Dutch: Netherlands Graduate School of Linguistics. PhD dissertation.
Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429–448.
Freudenthal, D., Pine, J., Jones, G., & Gobet, F. (2013). Flexible frames and the noun-verb asymmetry. In N. Sebanz, M. Knauf, M. Pauen, & E. I. Wachsmuth (Eds.). *Proceedings of the 35th annual meeting of the Cognitive Science Society* (pp. 2327–2332). Austin, TX: Cognitive Science Society.
Gil, D., & Tadmor, U. (2007). The MPI-EVA Jakarta child language database. A joint project of the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology and the Center for Language and Culture Studies, Atma Jaya Catholic University < http://lingweb.eva.mpg.de/jakarta/data_acquisition.php > .
Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436.
Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183–206.
Harris, Z. S. (1951). *Structural linguistics.* Chicago: The University of Chicago Press.
Haspelmath, M., Dryer, M.S., Gil, D., & Comrie, B. (2008). WALS online. Max Planck Digital Library.
Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy, 5*(3), 341–353.
Küntay, A. C., Koçbaş, D., & Taşçıı, S. S. (unpublished). Koç University Longitudinal Language Development Database on language acquisition of 8 children from 8 to 36 months of age.
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Mahwah, NJ: Lawrence Erlbaum Associates.
Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Vol. Ed.), *Children's language: Vol. 2*, (pp. 127–151). New York: Gardner Press.
Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91–117.
Mintz, T. H. (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 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.
Miyata, S. (2012). Japanese CHILDES: The 2012 CHILDES manual for Japanese < http://www2.aasa.ac.jp/people/smiyata/CHILDESmanual/chapter01.html > .
Miyata, S., & Nisisawa, H. Y. (2009). *MiiPro - Asato Corpus.* Pittsburgh, PA: Talkbank.
Miyata, S., & Nisisawa, H. Y. (2010). *MiiPro - Tomito Corpus.* Pittsburgh, PA: Talkbank.
Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*(4), 259–305.
Moran, S., Schikowski, R., Pajović, D., Hysi, C., & Stoll, S. (2016). The ACQDIV database: Min(d)ing the ambient language. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (Eds.), *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*. Paris, France: European Language Resources Association (ELRA).
Nazzi, T., Barrière, I., Goyet, L., Kresh, S., & Legendre, G. (2011). Tracking irregular morphophonological dependencies in natural language: Evidence from the acquisition of subject-verb agreement in French. *Cognition, 120*(1), 119–135.
Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology, 102*(4), 522–537.
Nisisawa, H. Y., & Miyata, S. (2009). *MiiPro - Nanami corpus.* Pittsburgh, PA: Talkbank.
Nisisawa, H. Y., & Miyata, S. (2010). *MiiPro - ArikaM corpus.* Pittsburgh, PA: Talkbank.
Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the Cognitive Science Society* (Vol. 26, pp. 1047–1052).
Payne, T. E. (1997). *Describing morphosyntax.* Cambridge: Cambridge University Press.
Pfeiler, B. (unpublished). Pfeiler Yucatec Child Language Corpus.
R Development Core Team (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing < http://www.r-project.org > .
Redington, M., Crater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425–469.
Rousseeuw, P. J., & Leonard, K. (1990). *Finding groups in data: An introduction to cluster analysis.* John Wiley & Sons.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
Sandoval, M., & Gómez, R. L. (2013). The development of nonadjacent dependency learning in natural and artificial languages. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(5), 511–522.
Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*(2), 105–134.
St Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition, 116*(3), 341–360.
Stoll, S., & Meyer, R. (unpublished). Audio-visional longitudinal corpus on the acquisition of Russian by 5 children.
Stoll, S., & Bickel, B. (2013). Capturing diversity in language acquisition research. In B. Bickel, L. A. Grenoble, D. A. Peterson, & A. Timberlake (Eds.). *Language typology and historical contingency: Studies in honor of Johanna Nichols* (pp. 195–260). Amsterdam: Benjamins.
Stoll, S., Lieven, E., Banjade, G., Bhatta, T. N., Gaenszle, M., Paudyal, N. P., …, Bickel, B. (unpublished). Audiovisual corpus on the acquisition of Chintang by six children.
Stumper, B., Bannard, C., Lieven, E., & Tomasello, M. (2011). "Frequent frames" in German child-directed speech: A limited cue to grammatical categories. *Cognitive Science, 35*(6), 1190–1205.
Van Kampen, A., Parmaksiz, G., van de Vijver, R., Höhle, B., Gavarró, A., & Freitas, M. J. (2008). Metrical and statistical cues for word segmentation: The use of vowel harmony and word stress as cues to word boundaries by 6- and 9-month-old Turkish learners. In *Language acquisition and development: Proceedings of GALA 2007* (pp.

313–324).
Wang, F. H., & Mintz, T. H. (2016). Statistical learning requires a two-step process. *Poster presented at the Boston University conference on language development (BUCLD)* Somerville, MA: Cascadilla Press.
Wang, H., Höhle, B., Ketrez, F. N., Küntay, A. C., Mintz, T. H., Danis, N., ... Sung, H. (2011). Cross-linguistic distributional analyses with frequent frames: The cases of German and Turkish. In N. Danis, K. Mesh, & H. Sung (Eds.). *Proceedings of 35th annual Boston University conference on language development (BUCLD)* (pp. 628–640). Somerville, MA: Cascadilla Press.
Weisleder, A., & Waxman, S. R. (2010). What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *Journal of Child Language, 37*(05), 1089–1108.
Widmer, M. (2015). *Die Rolle distributioneller Information im Spracherwerb: Kategorisierung mithilfe von häufigen Frames.* University of Zurich MA thesis.
Xiao, L., Cai, X., & Lee, T. H. (2006). The development of the verb category and verb argument structures in Mandarin-speaking children before two years of age. In Y. Otsu (Ed.). *Proceedings of the seventh Tokyo conference on psycholinguistics* (pp. 299–322). Tokyo: Hituzi Syobo.