BRIEF RESEARCH REPORT

What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English*

ADRIANA WEISLEDER

Stanford University

AND

SANDRA R. WAXMAN

Northwestern University

ABSTRACT

Recent analyses have revealed that child-directed speech contains distributional regularities that could, in principle, support young children's discovery of distinct grammatical categories (noun, verb, adjective). In particular, a distributional unit known as the FREQUENT FRAME appears to be especially informative (Mintz, 2003). However, analyses have focused almost exclusively on the distributional information available in English. Because languages differ considerably in how the grammatical forms are marked within utterances, the scarcity of cross-linguistic evidence represents an unfortunate gap. We therefore advance the developmental evidence by analyzing the distributional information available in frequent frames across two languages (Spanish and English), across sentence positions (phrase medial and phrase final), and across grammatical forms (noun, verb, adjective). We selected six parent–child corpora from the CHILDES database (three English; three Spanish), and analyzed the input when children were aged 2;6 or younger. In each language, frequent frames did indeed offer systematic cues to grammatical category assignment. We also identify differences in the accuracy of these frames across languages, sentences positions and grammatical classes.

To acquire a human language, children must not only learn individual words, but must also discover the distinct KINDS of words that are represented in their language (or grammatical categories, e.g. nouns, verbs, determiners) and how they map to meaning. Even within their first years, infants make significant headway in this arena. By age 0;9, they distinguish between two very broad kinds of words: content words (e.g. nouns, verbs, adjectives) vs. function words (e.g. determiners, prepositions) (Shi, Werker & Morgan, 1999). By age 1;1, they begin to make finer distinctions among the content words, teasing apart the grammatical form NOUN (e.g. 'cat') and mapping this form specifically to objects and object categories (e.g. cats). Over the next several months, they make finer distinctions still, teasing apart the forms ADJECTIVE and VERB, and mapping each to its associated range of meanings (properties and events, respectively) (Waxman & Lidz, 2006, for a review; Waxman & Booth, 2003).

How do infants accomplish this task? Because languages vary in the way that each grammatical form is marked in the 'surface' of the input (e.g. whether or not a modifier typically precedes or follows the noun, whether or not the nouns, adjectives and determiners are marked for grammatical gender), infants' accomplishments must rest upon an ability to glean grammatical form-class information from the surface structure of the utterances that they hear. But this claim – that the surface structure of a language provides reliable cues to grammatical category assignment – has been a controversial one. In the 1950s, structural linguists observed that words from the same grammatical category tend to appear in the same distributional environments (Harris, 1951). For example, words that appear with the morpheme *-ed* in English also tend to appear with the morpheme *-s*. Based on this observation, several researchers proposed that distributional regularities may serve as cues to grammatical form-class assignment (Maratsos, 1988; Maratsos & Chalkley, 1980; Kiss, 1973). However, this proposal was not endorsed universally. At issue was whether there is in fact sufficient structure in the language input to support the acquisition of grammatical categories and, if so, whether infants are in fact sensitive to the kinds of regularities present (Chomsky, 1965; Pinker, 1987).

For decades this challenge appeared insurmountable. In recent years, however, researchers using computational tools to examine the language input have documented that the distributional evidence available in natural-istic, child-directed speech may in fact offer strong cues to grammatical form-class assignment (Mintz, Newport & Bever, 2002; Redington, Chater & Finch, 1998; Cartwright & Brent, 1997). In a compelling recent demonstration, Mintz (2003) introduced the notion of FREQUENT FRAMES, a distributional pattern defined as two words that bracket one intervening word (e.g. *I __ you*). In an analysis of six corpora involving adult–child conversations, Mintz identified the most frequent frames in adults' speech.

(See (1–6) below for some representative examples.) He then considered whether within each such frame, the intervening words tended to belong to the same grammatical category. For some frames (e.g. (1–3)), the intervening words were predominantly verbs; for others (e.g. (4–6)), the intervening words were predominantly nouns. This suggests that frequent frames constitute a distributional unit that could, in principle, support the acquisition of distinct grammatical classes. Moreover, recent work reveals that young children and adults are sensitive to distributional patterns like these (Gómez, 2002; Gómez & Maye, 2005; Mintz, 2002). Taken together, these recent demonstrations have breathed new life into the hypothesis that distributional information in the input could support the discovery of distinct grammatical forms.

(1) I __ it
(2) you __ to
(3) I __ you
(4) the __ and
(5) a __ on
(6) the __ is

However, evidence to this effect has thus far focused primarily on the distributional information available in English. Because languages differ considerably in how the grammatical forms are marked on the surface of utterances, the scarcity of cross-linguistic evidence represents a gap that is important to fill. Consider, for example, free word order languages like Turkish, where the sequences in which words can appear vary freely. In such languages, the relevant distributional units may not be sequences of co-occurrences among words (as in the frequent frames for English); instead they may be sequences of co-occurrences among sublexical morphemes (Mintz, 2008). In principle, then, the distributional units that emerge as central in one language may differ from those that emerge in another.

Interestingly, we need not look as far as Turkish to appreciate the importance of cross-linguistic evidence. Even in languages more closely related to English, certain linguistic features may have consequences on the clarity and force of the distributional evidence for grammatical form classes. For example, Romance languages like French, Spanish and Italian exhibit considerable homophony among key function words. Although homophony is also present in English, it is primarily evident among content words, as opposed to function words. In cases involving content words, frequent frames may help to disambiguate among candidate meanings (e.g. *The brush is mine.* vs. *I brush your hair.*). In Spanish, however, homophony is not only evident among content words, but is also common among determiners (*la, los, las, una, unos, unas*) and other function words. For example, *los* can

function either as an article (e.g. *los niños juegan* 'the children play') or as an object pronoun (e.g. *Ana los quiere aquí* 'Ana wants them here'); *que* can mean either 'what' or 'that'; and *como* can mean either 'how', 'as' or 'eat'. In cases involving homophony among function words, the consequences for a frequent frames approach may be considerable: because function words are so frequent in the input, they often serve as framing elements in frequent frames. This particular type of homophony could therefore have adverse consequences on the clarity of distributional cues (Pinker, 1987; Cartwright & Brent, 1997). Chemla, Mintz, Bernal & Christophe (in press) recently reported that even in the face of this kind of homophony, the distributional evidence in a frequent frames analysis of French remained robust. However, because the Chemla *et al.* analysis considered the input to only one French-acquiring child, additional cross-linguistic evidence is clearly warranted.

In addition to homophony, there are other linguistic phenomena that could have consequences on the clarity and force of a distributional analysis. Consider, for example, a linguistic phenomenon known as noun dropping (Torrego, 1987; Snyder, Senghas & Inman, 2001). Noun dropping refers to a process by which nouns are omitted from the surface of a sentence when their meaning is recoverable from context (e.g. *Quiero una azul* 'I want a blue (one)', *El pequeño está dormido* 'The little (one) is sleeping'). Notice that noun dropping, which is more ubiquitous in some languages than others, is relevant to distributional analyses because, as a result of this syntactic process, nouns and adjectives often appear within the same frequent frames (e.g. following a determiner and preceding another element) (Waxman & Guasti, in press; Waxman, Senghas & Benveniste, 1997).

In view of these observations, our goal is to advance the cross-linguistic developmental evidence for distributional approaches in three key directions. First, we examine the distributional evidence available to young children acquiring Spanish and compare it to the evidence available to children acquiring English. Second, we consider the relative clarity of frequent frames for identifying the grammatical categories noun, verb and adjective. Third, we consider for the first time a new distributional environment. In addition to the frames described by Mintz (2003), in which two lexical items serve as framing elements (e.g. *you__ it*), we also consider phrase-final sequences in which the final utterance boundary serves as a framing element (e.g. *the__.*).

Our decision to consider these END-FRAMES was motivated by evidence that for young learners in particular, ends of utterances have a privileged status (Slobin, 1973). Infants and young children are sensitive to the prosodic cues that signal phrase boundaries (Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss & Kennedy, 1987). Thus, these prosodic cues might serve as framing elements. Moreover, in infant- and child-directed speech, key words

are often placed in utterance-final position and tend to receive exaggerated pitch peaks and increased durations (Fernald & Mazzie, 1991). Finally, children more successfully identify words presented in utterance-final, as compared to utterance-medial, position (Fernald & McRoberts, 1993; Shady & Gerken, 1999). Put succinctly, because infants and young children are attentive to phrase boundaries (and especially to those occurring in utterance-final position), if utterance boundaries constitute framing elements, and if the ends of phrases contain information that is relevant to grammatical form (Gleitman & Wanner, 1982; Morgan & Newport, 1981), then end-frames may constitute a potent source of information. Thus, by including both mid-frames (following Mintz, 2003) and end-frames, we have an opportunity to consider the accuracy of a frequent frames approach when the frames occur in different locations within the utterance (mid- and end-frames), and when they are bounded by different kinds of framing elements (words and utterance boundaries).

METHOD

Input corpora

We selected six parent–child corpora from the CHILDES database (MacWhinney, 2000); three in Spanish (Irene (Ojea, 2000), Koki (Montes, 1987) and María (López-Ornat, Fernández, Gallo & Mariscal, 1994)), and three in English (Eve (Brown, 1973), Naomi (Sachs, 1983) and Nina (Suppes, 1974)). We analyzed the utterances of the adult speakers in all sessions in which the child speaker was 2;6 or younger. The English corpora were among those previously examined by Mintz (2003). This ensured that our execution of the frequent frames analysis mirrored that reported by Mintz, and provided a point of comparison for Spanish.

Distributional analysis procedure

Our analytic procedure follows that of Mintz (2003).

Gathering the frames. Following Mintz (2003), we defined a FRAME as two linguistic elements with one word intervening. We considered two different types of frames. In the case of MID-FRAMES (**A__B**), the two framing elements were words (denoted by **A** and **B**), and the intervening word (denoted by __) varies. This is the frame analyzed by Mintz (2003). In the case of END-FRAMES (**A__.**), the first framing element was a word, the second an utterance-final boundary, and the intervening word varies. We segmented every adult utterance into three-element frames. For example, the utterance 'Look at the doggie over there' yielded five frames: four mid-frames ('look __ the', 'at __ doggie', 'the __ over', 'doggie __ there') and one end-frame ('over __.'). We did not include any frames that crossed an utterance boundary.

Selecting the frames. Next, we tabulated the frequency of each frame and selected for analysis the forty-five most frequent mid-frames and the forty-five most frequent end-frames.[1] These constitute our two groups of FREQUENT FRAMES.

Identifying the intervening words. We then listed all of the intervening words (both types (e.g. *dog, cat, run*) and tokens (e.g. every instance of *dog, cat* and *run*)) that appeared within each frequent frame; each list constituted a FRAME-BASED CATEGORY.

Identifying the grammatical category assignment of intervening words. We assigned each intervening word to a grammatical category (noun, verb, adjective, preposition, adverb, determiner, wh-word, conjunction or inter-jection). This step of the analysis was carried out by a native speaker of each language; a native Spanish speaker categorized all the words in the Spanish corpora and a native English speaker categorized those in the English corpora. To ensure that the same criteria were applied across the two languages, grammatical category assignments in each language were then checked by a Spanish–English bilingual. For any word that could be assigned to more than one grammatical category (e.g. in English, *party* is both a noun and verb; in Spanish, *vino* is both a noun and a verb), the corpus was consulted to identify the correct assignment. In every case, the corpus provided adequate evidence to assign the word's grammatical category.

Quantitative evaluation procedure

Our quantitative evaluation focuses on the accuracy, or consistency, of the frame-based categories. To compute 'Accuracy', we compared every pair of words that occurred within any frame (See Mintz (2003) and Redington *et al.* (1998) for full details). We identified each pair as either a 'Hit' (the two words were members of the same grammatical category) or a 'False Alarm' (the two words were members of different grammatical categories). We then calculated Accuracy by computing the proportion of Hits (Accuracy = Hits / (Hits + False Alarms)).

In the analyses reported here, we focus on Accuracy for token frequencies. However, it is important to point out that several different analyses yielded the very same pattern of results. First, independent analysis on word TYPES and TOKENS revealed the same pattern. Second, in addition to Accuracy, we also analyzed the 'Completeness' of the frame-based categories. This measure considers the proportion of word pairs from the same grammatical

---

[1] All of the frames in the analyzed set surpassed a frequency threshold of 5% in proportion to the total number of frames in each corpus. Moreover, when we restricted the analysis to the set of twenty-five most frequent frames we obtained comparable results, suggesting that the results are robust even under more restricted conditions.

category that were grouped together in the same frame (see Mintz (2003) for details). In the face of these strongly converging measures, we decided to report the results on Accuracy for token frequencies because this most closely reflects our goal of determining whether frequent frames are equally consistent across different languages and distributional environments. Finally, note that an analysis that strives for high Accuracy (even at the expense of Completeness) would result in categories with high internal consistency. Once this is achieved, some of the categories could be merged (based, for example, on their degree of overlap) in order to achieve higher Completeness (Mintz, 2003).

Baseline categorization: comparison to chance

To obtain a baseline categorization measure, we used a Monte Carlo method. Specifically, we computed accuracy scores for random word categories that were generated for each corpus. To accomplish this task, the intervening words within all of the frame-based categories were randomly distributed to form 'dummy' categories, which matched the frame-based categories in size. This random shuffling of the intervening words was repeated 1000 times, with accuracy computed on each shuffle. Accuracy scores obtained from these 1000 shuffles provided a baseline against which to compare the results from the frame-based categories and compute significance levels. For example, if only 2 out of 1000 shuffles matched the score obtained by the frame-based categorization method, the frame-based Accuracy score was said to be significantly above chance with a probability of 0·002.

RESULTS AND DISCUSSION

Table 1 offers an overview of the corpora for each child. Although there is within- and between-language variation in the size of the corpora, preliminary analyses ensured that there were no significant differences between the English and Spanish corpora in the total number of utterances analyzed, or in the number of types or tokens categorized by the frequent frames analysis (all p's > 0.3). Moreover, for every child, the words categorized by this analysis comprised a large fraction of the corpus (see Table 1, last column). This is important because it reveals that the frequent frames analysis 'captured' the words that predominate in the child's input.

Comparing frequent frames (actual) to baseline (chance)

For each corpus, we compared the Accuracy score for the frame-based categories to the corresponding baseline measure (derived using the Monte

TABLE 1. *Descriptive statistics for each corpus*

	Corpus	Number of utterances	Word tokens categorized	% Corpus categorized[a]	Word types categorized	% Corpus represented[b]
English	Eve	13,743	3,315	5·91%	393	64·89%
	Nina	14,478	6,296	8·58%	451	64·61%
	Naomi	7,148	1,814	6·24%	359	53·09%
	Mean	11,790	3,808	7%	401	61%
Spanish	Koki	4,585	815	0·51%	220	44·36%
	Irene	22,087	3,315	6·09%	393	75·68%
	María	10,916	2,079	4·35%	433	66·39%
	Mean	12,529	2,070	4%	349	62%

[a] # of tokens categorized in frequent frames/total # of tokens in the corpus.
[b] Percentage of tokens in the corpus whose *types* were categorized by frequent frames.
(e.g. If 'jacket' is categorized only once by frequent frames, but it appears 35 times in the corpus, the '% represented' would be 35 / total # of tokens in the corpus.)

TABLE 2. *Accuracy scores* (*token frequency*) *for each frame-type in each corpus*

		Mid-frames (A___B)	End-frames (A___.)
English	Eve	0·93	0·83
	(Baseline)	(0·48)	(0·40)
	Nina	0·98	0·92
	(Baseline)	(0·46)	(0·50)
	Naomi	0·96	0·85
	(Baseline)	(0·47)	(0·43)
	MEAN	**0·96**	**0·87**
	(Baseline)	**(0·47)**	**(0·44)**
Spanish	Koki	0·82	0·75
	(Baseline)	(0·24)	(0·38)
	Irene	0·68	0·52
	(Baseline)	(0·33)	(0·34)
	María	0·75	0·66
	(Baseline)	(0·34)	(0·33)
	MEAN	**0·75**	**0·64**
	(Baseline)	**(0·30)**	**(0·35)**

Carlo method described above) (see Table 2). The Accuracy scores for all corpora were significantly higher than baseline (all p's $<0·001$). This documents that, in the input for EACH CHILD, there is indeed consistent distributional information within frequent frames that converges on grammatical categories. This replicates Mintz (2003) and extends the work to a new frame-type (end-frames) and to a new language (Spanish).

TABLE 3. *Accuracy scores for each language, frame-type and grammatical class*

		Mid-frames	End-frames	Gram. class mean	Language mean
English	Noun	0·940	0·615	0·778	0·722*
	Verb	0·961	0·612	0·787	
	Adjective	0·723	0·488	0·601	
Spanish	Noun	0·643	0·659	0·651	0·592*
	Verb	0·824	0·646	0·735	
	Adjective	0·497	0·285	0·391	
Frame-type Mean		0·765**	0·551**		

* Significantly different from each other, $p < 0.05$.
** Significantly different from each other, $p < 0.001$.

Comparing across languages, frame-types and grammatical class

We next asked whether there were systematic differences in accuracy between the two languages, between the two types of distributional environments and between the grammatical classes. To address this question, we aggregated the frame-based categories for each language and frame-type[2] and calculated the accuracy score within each. We then categorized each frequent frame by the modal grammatical class of its intervening words (e.g. if the frame contained more nouns than words from any other grammatical class, it was classified as a Noun-frame). We submitted these accuracy scores to a three-way ANOVA: language (English vs. Spanish) by frame-type (mid-frame vs. end-frame) and by grammatical class (noun-frame vs. verb-frame vs. adjective-frame) (see Table 3). A main effect of language indicated that accuracy was higher in English ($M = 0.72$) than Spanish ($M = 0.60$) ($F(1,209) = 5.022$, $p = 0.026$, $\eta_p^2 = 0.023$). A main effect for frame-type indicated that accuracy was higher for mid-frames ($M = 0.77$) than for end-frames ($M = 0.55$) ($F(1,209) = 13.374$, $p < 0.001$, $\eta_p^2 = 0.06$). A main effect for grammatical class revealed higher accuracy for verb-frames ($M = 0.76$) than for noun-frames ($M = 0.71$), and higher accuracy for noun-frames than for adjective-frames ($M = 0.50$) ($F(2,209) = 5.154$, $p = 0.007$, $\eta_p^2 = 0.047$). All differences among these means were statistically reliable (Tukey's HSD, all p's < 0.05) (see Table 3). A subsequent analysis based on the data from each individual child's corpus ensured that these

[2] We note here there were eight English frames that contained only three to five word-types. These frames were excluded from further analysis so that they would not artificially inflate Accuracy. Their inclusion would only increase the size of our effect.

effects were not an artifact of the particular corpora we selected or of ag-
gregating data from different corpora.[3]

These findings provide support for the hypothesis 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. In particular, the analyses re-
ported here, coupled with a glance at Table 3, suggest that in both languages,
frequent frames contain robust cues for identifying the grammatical forms
noun and verb, but weaker cues for the form adjective. This outcome for
adjectives, although consistent with our proposal, should be interpreted with
some caution. Our analysis identified numerous noun- and verb-frames
in each language, and the accuracy of these frames tended to be high. By
contrast, we identified only a handful of adjective-frames (three and six, for
Spanish and English, respectively). Because these also included a number
of nouns and verbs, their accuracy was relatively low. Interestingly, and as
predicted, although this pattern was evident in both languages, a glance
at Table 3 suggests it was especially pronounced in Spanish.

GENERAL DISCUSSION

The current evidence provides support for the claim that the distributional
information in frequent frames contains cues that could be useful to young
children as they establish the main grammatical categories of their native
language. This work extends previous evidence in three ways. First,
it broadens the empirical base, examining the input available to children
acquiring Spanish as their mother tongue, and comparing it to the input
available to children acquiring English. We demonstrate that even in the
face of linguistic features that may render the distributional evidence in
Spanish less clear on the surface (including homophony and noun drop),
frequent frames nonetheless offer robust cues to grammatical form class.

Second, this work casts a wider distributional net, considering not only
frames that occur in utterance-medial positions and consist of only lexical
items as framing elements (mid-frames), but also frames that occur in
utterance final position and include utterance-final boundaries as framing

---

[3] There was a high degree of consistency among the frequent frames in the three corpora
in each language: 66% of the frequent frames were found in at least two of the three
corpora (52% of mid-frames, 79% of end-frames). Moreover, when we submitted the
accuracy scores from each corpus to a Language (English vs. Spanish) by Frame-type
(mid-frames vs. end-frames) ANOVA, we replicated the main effects of language and
frame-type. This provides further validation for aggregating over the corpora in each
language. In addition, Accuracy in our samples was not correlated with either the
number of utterances in the corpus, or the percentage of corpus categorized. This
provides assurances that the differences between English and Spanish reported here
cannot be attributed to either of these factors.

elements (end-frames). We demonstrate for the first time that in both English and Spanish, end-frames carry distributional cues to grammatical form class. Perhaps not surprisingly, the distributional information for end-frames (which, by definition, are less constrained by their surrounding elements than are mid-frames) is less accurate than that for mid-frames. Yet end-frame information, however noisy, may be especially useful to young children (Slobin, 1973), as they are better able to recognize words that occur at the ends of utterances (Fernald & McRoberts, 1993; Shady & Gerken, 1999).

Third, to the best of our knowledge, this is the first investigation to consider the relative accuracy of the distributional evidence available in frequent frames for the discovery of each of these grammatical categories. We found that frequent frames contained robust cues for the grammatical forms noun and verb, but weaker cues for adjectives, and that this pattern, although evident in both languages, appeared to be more pronounced in Spanish than in English.

On the whole, there were more commonalities than differences between the two languages, suggesting that this linguistic unit (the frame) stands as a powerful source of information for children acquiring either English or Spanish. At the same time, differences between the languages did emerge, which are likely due to linguistic features of the input. For example, in Spanish (as in other Romance languages) there is considerable homophony among function words (e.g. the word *la* can function either as a determiner (*la niña juega* 'the girl plays') or as a clitic (*ella la puso aquí* 'she put *it(f)* here')). These function words, which are highly frequent, often emerge as framing elements, and because they are homophonous, they affect the clarity of the distributional cues. Consider, for example, the mid-frame *la __ a*, which occurred frequently in Spanish. When *la* was used as a determiner, this frame picked out primarily nouns (e.g. *Lleva la muñeca a tu cuarto* 'Take the doll to your room'), but when *la* was used as a clitic, the very same frame picked out primarily verbs (e.g. *No la vuelvas a tocar* 'Don't *it-clitic* go to touch again'). Examples like this, considerably more common in Spanish than English, compromised frame accuracy.

Noun-drop also appears to have compromised accuracy in Spanish relative to English, especially in cases in which determiners emerged as framing elements. Consider, for example, the end-frame *un ___.*, in which the intervening words included nouns (e.g. *Dame un dulce* 'Give me a candy') and adjectives (e.g. *Eres un presumido* 'You are a vain (one)'). The distributional overlap between nouns and adjectives as a consequence of noun-drop was also evident, though less frequent, in mid-frames.

These observations suggest that certain features, including homophony and noun-drop, may indeed compromise the clarity of the distributional evidence available in frequent frames in Spanish. However, this is not to

say that as a whole, the distributional evidence available to Spanish-acquiring children is weaker, less informative or impoverished, relative to English. After all, children acquiring Spanish and English exhibit comparable developmental milestones. For example, between ages 1;9 and 2;0, infants learning either English or Spanish begin to distinguish adjectives from nouns, mapping the former primarily to object properties (e.g. color, texture) and the latter to object categories (e.g. dog, animal) (Waxman, Braun & Weisleder, in prep). Instead, we suggest that children acquiring different languages will rely on different kinds of distributional information to establish the grammatical categories of their language.

More specifically, the particular distributional cues that are most informative in one language may be different from those that are most informative in another. For example, in fixed word order languages like English, distributional evidence based on co-occurrences of words may be quite informative. But, in languages with rich inflectional morphology (e.g. Spanish, French, Italian), we suspect that additional cues to grammatical form class may be gleaned from distributions of morphemic, phonetic or prosodic properties. For example, our results suggest that differentiating between the roles of two homophonous words (particularly when they are function words) might be important for grammatical class assignment. Therefore one potential avenue for future research may be to explore whether there are differences in the phonetic or prosodic characteristics of homophonic pairs.

It will also be informative to focus on developmental matters. There is considerable evidence documenting that features of parental input vary not only across languages but also across development. In particular, the complexity of infant-directed speech changes over the first few years. How might these developmental changes affect the clarity of the distributional evidence? Perhaps the earliest input, like the input we have analyzed here, offers clearer distributional evidence to support the discovery of grammatical form classes. However, it is also possible that the early input, which is characterized by short utterances and exaggerated intonational contours (Fernald & Simon, 1984), can facilitate the identification of individual words in the continuous speech stream, but contains relatively little information to support the discovery of distinct grammatical forms.

It will also be important in future work to ascertain how frequent the frequent frames must be, and what proportion of the tokens available in the corpus they must capture, if they are to be informative. For example, in each of the corpora analyzed here, fewer than 10% of the tokens were categorized in the frequent frames analysis. Although this may seem, at first glance, to cast doubt on the utility of a frequent frames approach, a more careful examination reveals that these word types make up 44–76% of the tokens in the corpus (see Table 1, last column). Thus, those words that

were categorized constitute among the most commonly used words in the child's input, suggesting that frequent frames may offer information about grammatical categories for many of children's early words.

In sum, we have shown that there are distributional regularities in the linguistic input to children acquiring either English or Spanish that could, in principle, support the acquisition of distinct grammatical categories. Of course, because infants are sensitive to distributional regularities in domains other than language (Fiser & Aslin, 2002), the evidence presented here has broader implications for development. The challenge facing infancy researchers is to discover how infants make use of these regularities across development and across languages.

REFERENCES

Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press.

Cartwright, T. A. & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition* **63**(2), 121–70.

Chemla, E., Mintz, T. H., Bernal, S. & Christophe, A. (in press). Categorizing words using frequent frames: What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science*.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Boston, MA: MIT Press.

Fernald, A. & McRoberts, G. (1993). Effects of prosody and word position on infants' lexical comprehension. Paper presented at the Boston University Conference on Language Development, Boston.

Fernald, A. & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology* **27**(2), 209–221.

Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology* **20**, 104–113.

Fiser, J. & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences* **99**(24), 15822–26.

Gleitman, L. R. & Wanner, E. (1982). Language acquisition: The state of the art. In E. Wanner & L. R. Gleitman (eds), *Language acquisition: The state of the art*, 3–48. Cambridge: Cambridge University Press.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science* **13**(5), 431–36.

Gómez, R. L. & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy* **7**(2), 183–206.

Harris, Z. S. (1951). *Methods in structural linguistics*. Chicago, IL: University of Chicago Press.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B. & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition* **26**(3), 269–86.

Kiss, G. R. (1973). Grammatical word classes: A learning process and its simulation. *Psychology of Learning and Motivation* **7**, 1–41.

López-Ornat, S., Fernández, A., Gallo, P. & Mariscal, S. (1994). *La Adquisición de la Lengua Española* [The Acquisition of the Spanish Language], Madrid: Siglo XXI.

MacWhinney, B. (2000). *The CHILDES Project: Tools for anlyzing talk*, 3rd edn, vol. 2: The Database. Mahwah, NJ: Lawrence Erlbaum Associates.

Maratsos, M. (1988). The acquisition of formal word classes. In Y. Levy, I. M. Schlesinger & M. D. S. Braine (eds), *Categories and processes in language acquisition*, 21–44. Hillsdale, NJ: LEA.

Maratsos, M. P. & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. E. Nelson (ed.), *Children's language*, vol. 2, 127–214. New York: Gardner.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition* **30**, 678–86.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* **90**(1), 91–117.

Mintz, T. H. (2008). Distributional analysis as a method for categorizing words in infancy. Paper presented at the International Conference for Infant Studies, Vancouver, Canada.

Mintz, T. H., Newport, E. L. & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science* **26**(4), 393–424.

Montes, R. (1987). *Secuencias de clarificación en conversaciones con niños* [Clarification sequences in conversations with children]. *Morphe* 3–4: Universidad Autónoma de Puebla.

Morgan, J. L. & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior* **20**, 67–85.

Ojea, A. (2000). Llinàs-Ojea corpus: Irene. In MacWhinney, B. (ed.), The CHILDES database.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, 399–442. Hillsdale, NJ: Lawrence Erlbaum Associates.

Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science* **22**(4), 425–69.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent–child discourse. In K. E. Nelson (ed.), *Children's language*, 1–28. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shady, M. & Gerken, L. (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language* **2**, 163–75.

Shi, R., Werker, J. F. & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* **72**, B11–B21.

Slobin, D. I. (1973). Cognitive prerequisites for the acquisition of grammar. In C. A. Ferguson & D. I. Slobin (eds), *Studies of child language development*, 175–208. New York: Holt, Rinehart & Winston.

Snyder, W., Senghas, A. & Inman, K. (2001). Agreement morphology and the acquisition of Noun-drop in Spanish. *Language Acquisition* **9**(2), 157–73.

Suppes, P. (1974). The semantics of children's language. *American Psychologist* **29**, 103–114.

Torrego, E. (1987). Empty categories in nominals. Unpublished manuscript, University of Massachusets, Boston.

Waxman, S. R. & Booth, A. E. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science* **6**(2), 130–37.

Waxman, S. R., Braun, I. E. & Weisleder, A. (in prep) The breadth of adjective learning in English- and Spanish-acquiring infants.

Waxman, S. R. & Guasti, M. T. (in press). Linking nouns and adjectives to meaning: New evidence from Italian-speaking children. *Language Learning and Development*.

Waxman, S. R. & Lidz, J. (2006). Early word learning. In D. Kuhn & R. Siegler (eds), *Handbook of child psychology*, 6th edn, vol. 2, 299–335. Hoboken, NJ: Wiley.

Waxman, S. R., Senghas, A. & Benveniste, S. (1997). A cross-linguistic examination of the noun-category bias: Its existence and specificity in French- and Spanish-speaking preschool-aged children. *Cognitive Psychology* **32**(3), 183–218.

APPENDIX

Representative examples of noun-, verb- and adjective-frames for each language (English and Spanish) and frame-type (mid- and end-frames)

ENGLISH MID-FRAMES

the __ is (139 tokens, 78 types)
baby(2), bag(1), barn(1), basket(1), bear(2), blanket(1), book(1), bookcase(1), box(2), boy(3), brush(1), bug(3), bus(1), car(2), chicken(1), clip(1), cord(1), couch(1), cow(2), cream(1), crib(1), crumb(1), dinosaur(1), dog(4), doggie(3), doll(1), donkey(1), door(2), duck(4), egg(1), elbow(1), elephant(3), family(1), fish(2), floor(1), flower(1), fox(2), girl(2), gob(1), grass(1), head(1), horse(8), house(2), ice(1), juice(1), kangaroo(1), kite(1), kitty(3), lady(2), lamb(2), lipstick(1), lock(1), man(5), mommy(2), monkey(1), moon(6), mouse(3), neck(1), nest(1), nurse(1), ocean(1), paper(1), pig(2), powder(1), rabbit(4), radio(1), rest(1), roof(1), rooster(1), sleeper(1), smoke(2), squirrel(2), sun(6), this(1), tray(1), truck(4), window(1), zoo(1)

you __ the (446 tokens, 103 types)
arranging(1), at(1), ate(1), bite(1), boo(1), break(1), bring(7), broke(2), catch(3), chase(1), cleaning(1), close(6), closed(3), closing(3), cook(1), count(1), cover(1), crack(5), cracked(1), cracking(2), cut(1), cutting(4), do(1), draw(2), drawing(1), drink(1), drop(1), dropped(6), eat(1), eating(3), fed(4), feed(8), find(11), finish(1), fit(4), fix(3), for(3), found(2), get(14), give(6), giving(1), got(6), hang(1), have(3), hear(5), held(1), hide(1), hitting(1), hold(6), hugging(1), hurt(1), keep(1), knock(1), know(1), leave(1), let(1), like(27), loving(1), make(14), making(6), mean(1), moving(1), on(4), open(4), pat(2), patting(2), paying(1), pick(1), popped(1), pull(1), pulling(3), push(3), pushing(3), put(73), putting(21), read(9), reading(1), remember(4), roll(4), rolling(1), run(1), say(1), see(14), shut(2), smell(1), spilled(1), take(17), taking(4), tearing(2), think(4), throw(5), throwing(1), to(2), took(4), tore(1), turn(6), unscrew(1), unsticking(1), unwind(1), unwrapping(1), use(1), want(30), washing(1), with(4)

a __ one (50 tokens, 17 types)
big(5), black(2), blue(4), brown(2), chocolate(1), fine(1), grape(2), green(7), little(7), new(4), nice(3), nother(1), purple(5), red(2), touch(1), white(1), yellow(2)

SPANISH MID-FRAMES

la __ de (247 tokens, 121 types)
actuación$_n$(1), amiga$_n$(3), barba$_n$(1), basura$_n$(1), boca$_n$(1), bolsa$_n$(4), botella$_n$(1), busque$_n$(1), cabeza$_n$(2), caca$_n$(2), caja$_n$(1), cajita$_n$(1), cama$_n$(3),

15

cámara$_n$(1), camisa$_n$(1), camiseta$_n$(1), canción$_n$(20), cancioncilla$_n$(1), cantabas$_n$(1), cantidad$_n$(1), cara$_n$(15), carita$_n$(2), casa$_n$(9), casita$_n$(12), chaqueta$_n$(1), cinta$_n$(1), cola$_n$(1), colita$_n$(1), comida$_n$(3), cosita$_n$(1), cremallera$_n$(1), cuna$_n$(2), cunita$_n$(2), electricidad$_n$(1), encima$_{adv}$(1), era$_n$(3), espalda$_n$(3), esponja$_n$(2), esquina$_n$(1), falda$_n$(1), flor$_n$(2), foto$_n$(2), gorra$_n$(1), granja$_n$(1), habitación$_n$(4), hermana$_n$(2), historia$_n$(7), hoja$_n$(2), hojita$_n$(1), hora$_n$(2), jefa$_n$(1), llave$_n$(1), madrastra$_n$(3), madre$_n$(1), maestra$_n$(2), mamá$_n$(12), manina$_n$(1), manito$_n$(3), mano$_n$(11), manta$_n$(1), manzanilla$_n$(1), merienda$_n$(1), mesita$_n$(1), mitad$_n$(2), muevas$_v$(2), musiquita$_n$(3), naricina$_n$(1), naricita$_n$(1), nariz$_n$(1), novia$_n$(1), oreja$_n$(1), orejina$_n$(1), paciencia$_n$(1), papita$_n$(2), parte$_n$(2), patita$_n$(1), película$_n$(5), pelota$_n$(2), piel$_n$(1), pierna$_n$(2), pieza$_n$(1), plusvalía$_n$(1), posición$_n$(1), puerta$_n$(1), pulsera$_n$(1), resolución$_n$(1), revista$_n$(4), rota$_{adj}$(1), sabes$_v$(1), señora$_n$(4), sensación$_n$(1), silla$_n$(2), tapa$_n$(2), tarta$_n$(2), tenemos$_v$(1), tienda$_n$(8), trata$_v$(1), última$_{adj}$(1), uña$_n$(2), vaca$_n$(2), vana$_n$(1), vida$_n$(1), voz$_n$(3), zapatilla$_n$(1)

le __ a (169 tokens, 42 types)
aviso$_v$(1), cantamos$_v$(1), cogiste$_v$(1), cuentas$_v$(2), cuentes$_v$(3), da$_v$(1), dabas$_v$(1), das$_v$(1), decía$_v$(1), decías$_v$(1), dice$_v$(3), dices$_v$(10), dices$_v$(9), digas$_v$(1), dijiste$_v$(8), doy$_v$(1), enseñas$_v$(1), gu$_v$(1), gusta$_v$(3), gustaban$_v$(1), gustan$_v$(1), hace$_v$(1), haces$_v$(1), hiciste$_v$(3), iba$_v$(1), ibas$_v$(3), pasa$_v$(4), pasaba$_v$(2), paso$_v$(2), pediste$_v$(1), pegas$_v$(1), perdio$_v$(2), pone$_v$(1), quieres$_v$(1), regalamos$_v$(1), toca$_v$(5), va$_v$(14), vamos$_v$(16), van$_v$(1), vas$_v$(60), ve$_v$(1), voy$_v$(5)

un __ de (122 tokens, 37 types)
amigo$_n$(1), bastoncillo$_n$(1), beso$_n$(1), bibi$_n$(1), bocata$_n$(1), cachito$_n$(1), cartel$_n$(1), censo$_n$(1), chanchito$_n$(1), cuento$_n$(1), daikiri$_n$(1), dibujito$_n$(2), galito$_n$(1), gnomo$_n$(1), hueso$_n$(2), huevo$_n$(1), juego$_n$(2), librito$_n$(2), lunar$_n$(1), osito$_n$(1), paquete$_n$(2), par$_n$(2), pedacito$_n$(4), pellejito$_n$(1), pendiente$_n$(2), poco$_{adj}$(30), poquitito$_{adj}$(1), poquito$_{adj}$(45), pupuyu$_n$(1), rojo$_{adj}$(1), solomillo$_n$(1), sorbito$_n$(3), tipo$_n$(1), trozo$_n$(1), vaso$_n$(1), zapato$_n$(2)

ENGLISH END-FRAMES
that__. (378 tokens, 158 types)
afterwards(1), airplane(2), alright(3), animal(4), apart(1), awful(1), baby(3), be(1), becca(1), better(3), bibbie(1), blanket(2), block(1), blouse(1), blue(1), book(5), bottle(3), box(3), boy(2), button(3), cake(1), called(15), card(2), carton(1), cereal(1), chair(1), chirp(1), clay(1), cookie(1), corner(2), cute(1), d(1), daddy(1), darling(1), dog(3), doggy(2), dolly(3), door(1), easterbunny(2), either(1), eve(9), first(1), fit(3), flower(1), foot(2), for(2), frasers(1), frosting(1), fun(1), funny(3), game(2), girl(2), go(3), good(3), gun(1), hand(1), hard(1), hat(1), help(1), here(2), hole(3), home(3), honey(6), horse(2), hot(2), house(3), hurt(2), hurts(1), in(1), ink(1), is(18), it(4), jar(1),

16

jumped(1), kangaroo(1), kitty(1), knife(1), lamb(1), leila(1), letter(1), lion(1), long(1), makes(1), man(4), many(2), mean(1), means(2), mess(1), milk(1), mom(2), naomi(2), necessary(1), needs(1), nina(3), noise(6), nomi(24), off(2), okay(1), on(3), one(25), page(1), painful(1), papa(1), paper(2), part(3), pen(2), pencil(1), picture(13), piece(2), pillow(1), plate(2), please(1), pool(1), pot(1), present(2), pretty(1), puppet(1), puzzle(1), rabbit(2), racketyboom(2), rain(1), red(1), right(8), round(1), sarah(2), seat(1), shoe(2), side(2), song(1), sounds(1), soup(1), space(1), spell(1), spells(2), spoon(1), squeezes(1), stick(1), stool(1), story(2), stuff(1), sweetie(3), then(1), there(3), thing(2), tonight(1), too(2), tower(1), toy(1), train(2), tree(1), trip(1), two(1), valentine(2), water(1), way(14), what(1), window(2), you(1)

it__. (223 tokens, 152 types)
again(35), all(4), alone(7), along(2), anymore(3), apart(3), around(4), away(16), awhile(2), back(12), back(7), becca(1), before(1), belong(4), belongs(1), better(2), big(1), black(1), blue(3), break(2), bright(1), broke(3), broken(2), called(6), carefully(1), clean(1), cold(3), comes(3), cromer(1), cute(4), did(2), didn't(1), do(2), does(5), doesn't(1), doing(2), down(11), dried(1), drip(1), drop(1), either(3), eve(8), fantastic(1), feel(2), fell(1), first(6), fits(3), fixed(1), flies(1), fly(1), from(2), fun(8), funny(3), go(19), goes(7), going(3), goldie(1), gone(1), good(13), green(1), hard(4), here(11), hit(1), hmm(1), honey(12), horse(1), hot(2), hurt(6), hurts(3), in(24), into(2), is(115), isn't(1), itches(1), later(5), linda(1), look(1), louder(1), melted(1), moves(1), need(1), nina(1), no(2), nomi(27), nomi's(1), normally(1), now(4), off(40), okay(1), on(23), open(2), out(11), out(14), please(1), popped(1), pretty(1), rachel(1), raining(3), rattles(1), red(2), rest(1), right(2), rolled(1), rubs(1), sad(1), say(2), scary(1), shut(1), sideways(1), silly(1), sleeping(1), slowly(1), softly(1), somewhere(2), special(1), spin(1), stew(1), stop(1), sunny(1), sways(2), sweetheart(2), sweetie(1), swims(1), that(1), then(4), there(8), this(1), to(3), today(2), together(9), tomorrow(1), too(7), tore(1), untied(1), up(27), warm(1), was(5), wasn't(1), went(2), what(1), where(2), whistles(2), whole(1), will(3), with(2), work(1), works(2), would(1), yeah(4), yes(4), yet(3), yourself(4)

is__. (728 tokens, 120 types)
asleep(1), awake(1), baby(2), barking(1), better(1), bigger(2), billowing(1), black(1), blue(7), broken(6), called(7), clipclop(1), cold(3), colleen(1), coming(1), cool(2), cromer(2), crying(1), dancing(1), different(1), doing(1), dolly(3), drawing(1), eating(4), empty(2), eve(7), evecummings(1), fine(1), finished(1), fraser(2), frasers(1), fred(1), froggy(1), frosty(1), funny(1), furniture(1), good(1), grass(1), green(1), happening(1), hard(1), he(37), heavy(1), here(3), home(1), honey(4), hot(5), humm(1), hungry(1), indeed(2), inside(1), it(179), jackie(1), jenko(1), jim(1), jumping(1), kanga(1), left(1),

17

lemon(2), light(1), like(1), little(1), locked(1), missing(1), mommys(1),
next(1), nice(2), nina(2), nomi(9), now(1), nuzzling(1), ohio(3), ok(1), one(1),
open(1), out(3), outside(1), painting(1), papa(5), racketyboom(1), raking(1),
reading(1), ready(1), red(4), ricci(1), right(1), roo(1), running(1), saying(1),
she(17), sleeping(6), smaller(1), snoopy(1), soap(1), soup(1), stuck(3),
sugar(1), tea(1), that(212), there(4), this(80), time(1), timmy(1), tired(3),
today(1), too(1), upsidedown(1), walking(1), weak(1), wearing(1), what(4),
where(1), white(1), who(1), woopsie(1), working(1), yawning(1), yellow(4),
yes(1), yours(1)

SPANISH END-FRAMES

los __. (369 tokens, 171 types)

$abrimos_v$(1), $amiguetes_n$(1), $ángeles_n$(1), $animales_n$(1), $anises_n$(1),
$autobuses_n$(1), $bambis_n$(2), $barcos_n$(2), $barriletes_n$(3), $bebes_n$(3), $bebitos_n$(1),
$besos_n$(1), $bibis_n$(2), $bichos_n$(1), $bocatas_n$(1), $bracitos_n$(3), $brazos_n$(2),
$caballitos_n$(1), $caballos_n$(2), $cables_n$(1), $cacharritos_n$(1), $cacharros_n$(1),
$calcetines_n$(3), $carros_n$(2), $catalanes_n$(1), $cereales_n$(1), $chicos_n$(1),
$chiquitos_n$(1), $ciervos_n$(1), $coches_n$(2), $colores_n$(1), $columpios_n$(5), $comen_v$(1),
$como_v$(1), $conejitos_n$(1), $conejos_n$(2), $conoce_v$(2), $contamos_v$(2), $cordones_n$(1),
$cristales_n$(1), $cucharros_n$(1), $cuento_v$(1), $cuentos_n$(3), de_{prep}(3), $deditos_n$(3),
$dedos_n$(1), $demás_{adj}$(13), $descuidas_v$(1), $diablitos_n$(1), $días_n$(2), $dibujitos_n$(1),
$dibujos_n$(1), $dientes_n$(7), dos_{adj}(9), $elefantes_n$(2), $elotes_n$(1), $elotitos_n$(1),
$enemigos_n$(1), $enseño_v$(1), $fantasmas_n$(1), $fideos_n$(1), $gallegos_n$(2), $gallos_n$(1),
$ganó_v$(1), $gato_n$(1), $guardamos_v$(1), $guardes_v$(1), $guardo_v$(1), $helados_n$(1),
$hiciste_v$(1), $honguitos_n$(6), $huesos_n$(1), $jamases_n$(1), $juguetes_n$(10), $lavo_v$(1),
$leotardos_n$(2), $libritos_n$(5), $mayores_{adj}$(1), $meto_v$(1), $mezclamos_v$(2),
$mickeys_n$(1), $mimitos_n$(1), $mocos_n$(2), $moquetes_n$(1), $moquitos_n$(1),
$moros_n$(2), $mosqueteros_n$(2), $mosquitos_n$(1), $muebles_n$(2), $muñecos_n$(3),
$muñequitos_n$(4), $nenes_n$(12), $nervios_n$(2), $niñitos_n$(1), $niños_n$(15),
$números_n$(3), $ojinos_n$(3), $ojitos_n$(5), $ojos_n$(8), $ositos_n$(2), $osos_n$(3), $otros_{adj}$(3),
$pajarinos_n$(2), $pajaritos_n$(1), $palitos_n$(1), $palotes_n$(1), $pantalones_n$(4),
$papás_n$(3), $papelitos_n$(1), $pasajeros_n$(1), $patitos_n$(5), $patos_n$(2), $peces_n$(3),
$pedacitos_n$(1), $pedales_n$(2), $pegotitos_n$(1), $pellejitos_n$(1), $pelos_n$(1),
$pequeñitos_{adj}$(1), $perdio_v$(1), $perritos_n$(7), $perros_n$(1), $pescaditos_n$(4),
$petes_n$(1), $piececitos_n$(1), $pies_n$(15), $pifufos_n$(5), $pisos_n$(1), $pollitos_n$(2),
$ponemos_v$(1), $pongas_v$(1), $poquís_n$(1), $portales_n$(1), $primos_n$(1), que_{pron}(1),
$quesitos_n$(1), $quiere_v$(1), $quieres_v$(1), $quillitos_n$(2), $recoges_v$(1), $regalamos_v$(1),
$regaló_v$(1), $regalo_v$(1), $regalos_n$(1), $reyesmagos_n$(7), $sacó_v$(2), $saltitos_n$(1),
$sapos_n$(1), $seco_v$(2), $semaforos_n$(1), $señores_n$(4), $silloncillos_n$(2), $techos_n$(1),
$tiene_v$(1), $titos_n$(1), $toquen_v$(1), $tumbamos_v$(1), $tuyos_{pron}$(4), $vagones_n$(7),
$varones_n$(1), $vecinos_n$(1), $verá_v$(1), ves_v(1), $vestidos_n$(1), $videos_n$(1), vio_v(1),
$viris_v$(1), $zapatitos_n$(6), $zapatos_n$(8)

te __. (310 tokens, 134 types)

acalores$_v$(1), acercas$_v$(1), aclaras$_v$(2), acuerdas$_v$(39), advierto$_v$(1), ahogas$_v$(2), alejes$_v$(1), apetece$_v$(3), apuestas$_v$(2), atreves$_v$(1), ayudo$_v$(2), bajamos$_v$(1), bañamos$_v$(1), bañas$_v$(1), bañaste$_v$(2), busco$_v$(1), caes$_v$(4), caigan$_v$(1), caigas$_v$(1), cante$_v$(1), cantemos$_v$(1), cas$_v$(1), castiga$_v$(1), coja$_v$(1), conozco$_v$(1), costipas$_v$(1), cuente$_v$(2), cuento$_v$(1), da$_v$(2), decia(1), dejo(2), dice(4), dicen(2), diga(1), digo(13), dija(1), dijo(4), dio(1), diste(4), dolIa(1), doy(1), duele$_v$(3), eh$_{int}$(1), enfadabas$_v$(1), enfadas$_v$(3), enfadruscas$_v$(1), enseño$_v$(1), ensucias$_v$(1), entendí$_v$(1), enteras$_v$(1), entiende$_v$(1), entiendo$_v$(7), escapes$_v$(1), escondas$_v$(4), escuche$_v$(1), escucho$_v$(1), escurres$_v$(1), explico$_v$(1), expliques$_v$(1), gusta$_v$(27), gustan$_v$(4), gusto$_v$(3), hace$_v$(5), haga$_v$(1), hago$_v$(1), hizo$_v$(1), hundes$_v$(1), importa$_v$(2), lavan$_v$(1), levantas$_v$(1), levantaste$_v$(1), llamas$_v$(6), lleva$_v$(1), manchas$_v$(3), manches$_v$(2), mates$_v$(1), mato$_v$(1), moja$_v$(2), mojaste$_v$(1), mojes$_v$(1), molesta$_v$(2), moleste$_v$(1), muerda$_v$(1), ocurra$_v$(1), oí$_v$(2), oigo$_v$(1), oye$_v$(1), parece$_v$(15), parezca$_v$(1), pasa$_v$(11), pasaba$_v$(1), paso$_v$(1), pega$_v$(1), pegas$_v$(1), peinamos$_v$(1), peino$_v$(2), perdono$_v$(1), pinchar$_v$(1), pone$_v$(3), pongo$_v$(1), preocupes$_v$(2), puso$_v$(1), qué$_{wh}$(1), quemo$_v$(1), quiero$_v$(1), quito$_v$(1), quitó$_v$(1), rebelaste$_v$(1), ries$_v$(2), riño$_v$(1), sabes$_v$(2), sabrías$_v$(1), saco$_v$(1), sale$_v$(1), salió$_v$(1), seca$_v$(2), seco$_v$(1), seque$_v$(2), sequemos$_v$(1), sientas$_v$(3), sientes$_v$(1), tiraste$_v$(1), va$_v$(1), vas$_v$(2), vayas$_v$(1), ve$_v$(6), veamos$_v$(1), veía$_v$(1), ven$_v$(1), veo$_v$(4), vestimos$_v$(2), viste$_v$(2)

está __. (287 tokens, 158 types)

abajo$_{adv}$(1), abierto$_{adj}$(1), acá$_{adv}$(1), acabando$_v$(2), adónde$_{wh}$(1), agarrado$_v$(1), agusto$_{adj}$(1), ahí$_{adv}$(19), alto$_{adj}$(1), angelines$_n$(1), apagada$_{adj}$(1), apísima$_{adj}$(1), apuntando$_v$(1), aquía$_{adv}$(14), arregladita$_{adj}$(1), arregladito$_{adj}$(1), así$_{adv}$(1), bailando$_v$(1), bañando$_v$(5), batido$_{adj}$(1), bien$_{adv}$(18), bonito$_{adj}$(2), buenísimo$_{adj}$(1), bueno$_{adj}$(5), cabeza$_n$(1), calentito$_{adj}$(1), caliente$_{adj}$(6), cansada$_{adj}$(1), cansadito$_{adj}$(1), cariño$_n$(1), carlitos$_n$(2), carne$_n$(1), carnina$_n$(1), cerrada$_{adj}$(1), cerrado$_{adj}$(2), chica$_{adj}$(1), chupando$_v$(1), colonia$_n$(1), columpiando$_v$(1), comiendo$_v$(1), cuqui$_n$(1), dentro$_{adv}$(1), descalzada$_{adj}$(1), descalzo$_{adj}$(1), descansando$_v$(2), diciendo$_v$(3), disfrazado$_{adj}$(1), distraído$_{adj}$(1), donde$_{con}$(1), dormida$_{adj}$(1), dormido$_{adj}$(1), dura$_{adj}$(1), durmiendo$_v$(2), duro$_{adj}$(1), eh$_{int}$(3), en$_{prep}$(1), enfermita$_{adj}$(1), enseñasela$_v$(1), es$_v$(1), escuchando$_v$(1), esponja$_n$(1), esta$_n$(3), estornudando$_v$(1), fea$_{adj}$(2), fría$_{adj}$(1), frio$_{adj}$(1), grabando$_v$(1), gracias$_{int}$(1), gritando$_v$(2), grover$_n$(3), guapa$_{adj}$(4), guapo$_{adj}$(1), guardada$_{adj}$(1), guardadita$_{adj}$(1), gustando$_v$(1), hablando$_v$(1), haciendo$_v$(25), hay$_v$(1), hija$_n$(3), igual$_{adj}$(1), irene$_n$(3), jugando$_v$(4), koki$_n$(3), la$_{det}$(1), ladrando$_v$(1), limpio$_{adj}$(1), limpita$_{adj}$(3), limpito$_{adj}$(1), linda$_{adj}$(1), llorando$_v$(5), lloviendo$_v$(4), luis$_n$(2), mal$_{adv}$(1), mala$_{adj}$(1), malito$_{adj}$(1), mamá$_n$(2), mami$_n$(1), mano$_n$(1), maría$_n$(4), masticadin$_{adj}$(1), mejor$_{adj}$(1), merendando$_v$(1), mintiendo$_v$(1), mirando$_v$(1),

mojado$_{adj}$(1), nada$_n$(1), nena$_n$(2), niña$_n$(5), no$_{adv}$(2), ocupado$_{adj}$(1), oscuro$_{adj}$(2), otra$_{adj}$(2), paco$_n$(1), papá$_n$(5), pedrin$_n$(1), peque$_{adj}$(1), pequeño$_{adj}$(2), perrita$_n$(1), peterpan$_n$(3), pieza$_n$(1), prohibido$_{adj}$(1), pulsera$_n$(1), qué$_{wh}$(7), quedando$_v$(2), quién$_{wh}$(1), quilli$_n$(1), quique$_n$(1), retorcido$_{adj}$(1), revista$_n$(1), rica$_{adj}$(4), rico$_{adj}$(12), rota$_{adj}$(2), rotita$_{adj}$(1), rotito$_{adj}$(1), roto$_{adj}$(2), sacando$_v$(1), santi$_n$(1), seco$_{adj}$(1), sentado$_{adj}$(3), solito$_{adj}$(2), sucio$_{adj}$(2), suficiente$_{adj}$(1), tampoco$_{adv}$(1), tarde$_{adv}$(4), terminamos$_v$(1), tito$_n$(1), tocando$_v$(1), todo$_{adj}$(3), tolumpiando$_v$(1), triste$_{adj}$(1), usado$_{adj}$(1), vacilando$_v$(1), vale$_{int}$(1), verdad$_{int}$(2), vestida$_{adj}$(1)