Cognitive Science: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653634

Lexical Categories at the Edge of the Word

Luca Onnis [a]; Morten H. Christiansen [a]
[a] Department of Psychology, Cornell University,

Online Publication Date: 01 January 2008

To cite this Article Onnis, Luca and Christiansen, Morten H.(2008)'Lexical Categories at the Edge of the Word',Cognitive Science: A Multidisciplinary Journal,32:1,184 — 221
To link to this Article: DOI: 10.1080/03640210701703691
URL: http://dx.doi.org/10.1080/03640210701703691

Cognitive Science 32 (2008) 184–221

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1080/03640210701703691

Lexical Categories at the Edge of the Word

Luca Onnis, Morten H. Christiansen

Department of Psychology, Cornell University

Received 18 November 2005; received in revised form 12 September 2006; accepted 18 September 2006

Abstract

Language acquisition may be one of the most difficult tasks that children face during development. They have to segment words from fluent speech, figure out the meanings of these words, and discover the syntactic constraints for joining them together into meaningful sentences. Over the past couple of decades, computational modeling has emerged as a new paradigm for gaining insights into the mechanisms by which children may accomplish these feats. Unfortunately, many of these models assume a computational complexity and linguistic knowledge likely to be beyond the abilities of developing young children. This article shows that, using simple statistical procedures, significant correlations exist between the beginnings and endings of a word and its lexical category in English, Dutch, French, and Japanese. Therefore, phonetic information can contribute to individuating higher level structural properties of these languages. This article also presents a simple 2-layer connectionist model that, once trained with an initial small sample of words labeled for lexical category, can infer the lexical category of a large proportion of novel words using only word-edge phonological information, namely the first and last phoneme of a word. The results suggest that simple procedures combined with phonetic information perceptually available to children provide solid scaffolding for emerging lexical categories in language development.

Keywords: Phonological bootstrapping; Lexical categories; Computational models; Language acquisition; Cross-linguistic corpus analyses; Statistical learning; Neural networks

1. Introduction

By their third year of life children have already learned a great deal about how words are combined to form complex sentences. This achievement is particularly puzzling for cognitive science for at least three reasons: First, whatever learning mechanisms children bring to bear, they are thought to be of simpler computational complexity than adults'; second, children acquire most syntactic knowledge with little or no direct instruction; third, learning the

Correspondence should be addressed to Luca Onnis, Department of Psychology, Cornell University, 245 Uris Hall, Ithaca, NY 14850. E-mail: lo35@cornell.edu

complexities of linguistic structure from mere exposure to streams of sounds seems vastly complex and unattainable.

A particularly hard case is the discovery of lexical categories such as nouns and verbs, without which adult linguistic competence cannot be achieved. Indeed the very core of syntactic knowledge is typically characterized by constraints governing the relationship between lexical categories of words in a sentence. But, acquiring this knowledge presents the child with a "chicken-and-egg" problem: The syntactic constraints presuppose the grammatical categories in terms of which they are defined; and the validity of grammatical categories depends on how far they support syntactic constraints. Given the importance of this knowledge in language acquisition much debate has centered on how grammatical category information is gleaned from raw input. Even assuming that the categories themselves are innate (e.g., Pinker, 1984), the complex task of assigning lexical items from a specific language to such categories must be learned (e.g., the sound /su/ is a noun in French [*sou*] but a verb in English [*sue*]). Crucially, children still have to map the right sound strings onto the right grammatical categories while determining the specific syntactic relations between these categories in their native language.

In trying to explain how linguistic knowledge develops, the field of language acquisition has recently benefited from a wave of computational modeling and a series of large-scale statistical analyses derived from samples of natural language spoken to children. Rising above decades of skepticism motivated by poverty of the stimulus arguments, and overcoming the limitations of slow computers and small datasets, these large-scale analyses have started to provide evidence that natural languages may be abundant and redundant with statistical cues to their structure. The idea behind these analyses is to provide a statistical estimate, using simple statistical methods, of how far a system could get into discovering linguistic structure if a particular source of information was used. Some results immediately appeared rather surprising when they were first published. For instance, Redington, Chater, and Finch (1998) showed that a clustering algorithm produced clusters of words quite close to actual syntactic categories based on a simple source of information, namely the words that immediately preceded or followed a target word (e.g., in "*The igatu is here*" the lexical category noun for the unknown word *igatu* may be potentially gleaned from the facts that nouns typically precede *the* and follow *is*). These analyses have promoted a new way of looking at language acquisition: Many cues to higher level linguistic representations may actually be low-level or simple surface features of languages (such as the distributional information in the example above), which until recently had been dismissed a priori as largely uninformative. Moreover, it may be the case, as we shall argue in this article, that phonetic cues readily available to young infants could actually inform the discovery of higher-level properties of language.

Our first aim in this article is to make a further contribution to the study of probabilistic cues to language acquisition by assessing a potential source of information that has not been evaluated before, namely word beginnings and endings. In a first experiment we evaluated the usefulness of morphological markers, prefixes and affixes, for discovering lexical categories in English, as proposed by Maratsos and Chalkley (1980).

A second related aim in this article is to incorporate assumptions that make language modeling more plausible from a developmental perspective. Especially in the early stages of language development many cues to language structure may be *useful* in theory, but not necessarily *usable,* because they require an already sophisticated linguistic system to be in

place. For example, although morphology turns out to be potentially useful for learning about lexical categories, it seems unlikely that the necessary knowledge of morphology is in place in the second year of life when lexical units start to emerge. Given that morphological cues thus are less likely to play a role in the initial states of language discovery, is there an equivalent source of information that requires minimal linguistic assumptions, and which would be more *usable* than morphology? We propose that word edges (i.e., the first and last phoneme of words) are as useful as morphological markers in that they provide reliable scaffolding to develop the first rudiments of lexical categories. More important, word edges are also likely to be more usable for learning about lexical categories because they can be used without knowing anything about morphology. Thus, the plausibility of a particular cue depends not only on the presence in the speech signal of information relevant to syntactic organization but also on the capacities of learners to pick up these cues (see also Jusczyk, 1999).

We propose that plausible language models of lexical categorization also need to comply with two additional criteria, one being that the learning mechanism should be successful at generalizing to unseen lexical items; in particular in our case good generalization rests on the ability to classify novel words for which the lexical category is not available. We provide initial evidence that word-edge segments can be used successfully to determine the broad lexical category of a novel word. As a second criterion, we suggest that models should be applied not only within a single language but also across different languages. We therefore extended the analyses on English to languages that progressively differ from English, namely Dutch, French, and Japanese.

The plan of the article is as follows: We first briefly review the developmental literature on learning lexical categories in early language acquisition. We then start by estimating the usefulness of morphological affixes—prefixes such as *re-* in *reuse* and suffixes like *-al* in *magical*, in Experiment 1. Morphological markers have been proposed to assist lexical categorization in English (Maratsos & Chalkley, 1980), although there are no studies that assess empirically their contribution. In Experiment 2, we note that even though this source of information is potentially available in the input, children are not spoon-fed a list of morphological prefixes and suffixes. However, empirical evidence suggests that infants do pay particular attention to the beginnings and endings of words. Hence, in Experiment 2, we suggest that a more psychologically plausible mechanism is one that learns to categorize words based on word-edge segments, requiring no *a priori* knowledge of morphology. Experiments 1 and 2 rest on a largely supervised model—discriminant analysis—for which the lexical category labels are given. Therefore, these experiments show that there are significant correlations between the beginnings and endings of words and their respective lexical categories, suggesting their potential usefulness in supporting the learning of such categories, but implying that some knowledge of the lexical categories must come from some other source. In Experiment 3 we test a largely unsupervised model, single-layer perceptrons trained on a small subset of words for which the word-edge-to-category mapping is known, and we show that these simple learning models can assign correct lexical labels to a large number of novel untagged words. In Experiments 4, 5, and 6 we show that our simple word-edge procedure extends well to lexical categorization of languages that are progressively more distant from English, namely Dutch (a Germanic language), French (a Romance language with heavier but also more ambiguous inflection), and Japanese (a non Indo-European language). To end with, we discuss in detail

the distinction between the *usefulness* and the *usability* of cues. We argue that although many cues to linguistic structure are potentially *useful* in the input, and can be assessed computationally, some may be perceptually and cognitively available to language learners earlier than others, and hence may be more *usable* in building a first rudimentary knowledge of grammar.

2. Where does information for lexical categories come from?

There are three major sources of information that children could potentially bring to bear on solving the problem of inducing lexical categories: innate knowledge in the form of linguistic universals (e.g., Pinker, 1984); language-external information (e.g., Bowerman, 1973), concerning observed relationships between language and the world; and language-internal information, such as aspects of phonological, prosodic, and distributional patterns that indicate the relation of various parts of language to each other (e.g., Morgan & Demuth, 1996). Although not the only source of information involved in language acquisition, we suggest that probabilistic language-internal information may guide the child into syntactic development. The key idea is that learners may use information present in the speech signal to gain valuable knowledge about the syntactic organization of their native language. Computational models are particularly apt at investigating language-internal information because it is now possible to access large computerized databases of infant-directed speech and quantify the usefulness of given internal properties of a language.

Several studies have already assessed the usefulness of distributional, phonological, and prosodic cues. Distributional cues refer to the fact that lexical items in the speech stream tend to follow specific relations of co-occurrence. For instance, determiners typically precede nouns, but do not follow them (*the car*/**car the*). Corpus analyses have demonstrated that distributional patterns of word co-occurrence give useful cues to grammatical categories in child-directed speech (e.g., Finch, Chater, & Redington, 1995; Mintz, 2003; Monaghan, Chater, & Christiansen, 2005; Redington et al., 1998). Given that function words like articles and prepositions primarily occur at phrase boundaries (e.g., initially in English and French; finally in Japanese) they may also reveal syntactic structure. This is confirmed by corpus analyses (Mintz, Newport, & Bever, 2002) and results from artificial language learning (Green, 1979; Morgan, Meier, & Newport, 1987; Valian & Coulson, 1988).

Prosodic cues for word and phrasal/clausal segmentation may help uncover syntactic structure (e.g., Gerken, Jusczyk, & Mandel, 1994; Gleitman & Wanner, 1982; Kemler-Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989; Morgan, 1996). Differences in pause length, vowel duration, and pitch often align with phrase boundaries in both English and Japanese child-directed speech (Fisher & Tokura, 1996). Infants seem highly sensitive to such language-specific prosodic patterns (Gerken et al., 1994; Kemler-Nelson et al., 1989; for reviews, see Gerken, 1996; Jusczyk & Kemler-Nelson, 1996; Morgan, 1996)—a sensitivity that may start *in utero* (Mehler et al., 1988). Prosodic information also improves sentence comprehension in 2-year-olds (Shady & Gerken, 1999). Results from artificial language learning experiments with adults show that prosodic marking of syntactic phrase boundaries facilitates learning (Morgan et al., 1987; Valian & Levitt, 1996). Evidence from event-related brainwave potentials in adults showing that prosodic information has an immediate effect on syntactic

processing (Steinhauer, Alter, & Friederici, 1999) further underscores the importance of this cue.

Finally, phonological cues have also been shown to be useful for grammatical acquisition. For instance, adults are sensitive to the fact that English disyllabic nouns tend to receive initial-syllable (trochaic) stress whereas disyllabic verbs tend to receive final-syllable (iambic) stress (Kelly, 1988) and such information is also present in child-directed speech (Monaghan et al., 2005). Detailed acoustic analyses have shown that even noun-verb ambiguous disyllabic words that change grammatical category but not stress placement can be differentiated by syllable duration and amplitude differences (Sereno & Jongman, 1995). Moreover, both lexical access and on-line sentence comprehension is influenced by how typical nouns and verbs sound with respect to other words in the same lexical category (Farmer, Christiansen, & Monaghan, 2006). Experiments indicate that children as young as 3 years old are sensitive to differences in number of syllables, even though few multisyllabic verbs occur in child-directed speech (Cassidy & Kelly, 1991, 2001). Other phonological cues—including stress, vowel quality, and duration—may help distinguish content words (nouns, verbs, adjectives, and adverbs) from function words (e.g., determiners, prepositions, conjunctions) in English (e.g., Cutler, 1993; Gleitman & Wanner, 1982; Monaghan et al., 2005; Morgan, Shi, & Allopenna, 1996; Shi, Morgan & Allopenna, 1998).

We have briefly reviewed literature that suggests that several probabilistic cues internal to the language may assist the emergence of linguistic knowledge, in particular lexical categories. In Experiment 1, we assess the usefulness of another potential source of information, namely morphological marking. Morphology has attracted a large interest in language research since the early days of modern linguistics. In the tradition of structuralist linguistics, Wells (1947) suggested that morpheme classes might be the same as syntactic categories, and Z. Harris (1946) proposed procedures for discovering lexical categories from morphemes in sequences of words. Morphological analysis is nowadays at the core of many efforts in computational linguistics, where part-of-speech taggers and syntactic parsers make successful use of morphemic information (e.g., Nagata, 1999) and word edges, namely the first and last letter of words, in written text (Mikheev, 1997). The issue at stake in these models is to develop simple procedures that are not labor intensive in order to tag large amounts of unknown words in new texts with lexical category information. Efficient procedures are "portable" (i.e., they can be applied to the tagging of different languages). Therefore, the procedure developed in this study might appeal to the computational linguistics community as well.

3. Experiment 1: Morphological cues in grammatical categorization

Intuitively, morphological patterns across words seem informative for grammatical categorization. For instance, Maratsos and Chalkley (1980) noted that English words that are observed to have both *-ed* and *-s* endings are likely to be verbs. Artificial language learning results show that adults, children, and infants are better at learning grammatical categories cued by word internal patterns (Brooks, Braine, Catalano, & Brody, 1993; Frigo & McDonald, 1998; Gerken, Wilson, & Lewis, 2005). Other artificial language learning experiments indicate that duplication of morphological patterns across phrase-related items (e.g., concord morphology

in Spanish: *Los Estados Unidos*) facilitates learning (Meier & Bower, 1986; Morgan et al., 1987), and diminutives assist the learning of gender in Russian (Kempe & Brooks, 2005). Besides suffixes, children may also exploit prefixes, although to our knowledge little work has been done to assess empirically the usefulness of this cue. Despite their apparent usefulness, some cues are ambiguous with regard to lexical assignment. For instance, *-s* can serve as a third person singular inflection on a verb or a plural marker on countable nouns. In general, English has relatively little inflection, much less than most Indo-European languages. In addition, unlike many other European languages, English nouns do not carry gender information. Furthermore, many prefixes and affixes are historical borrowings from Latin, more likely to be used in Latinate words, which tend to be rare and low-frequency in child-directed speech. Given these considerations, is there a way to estimate the usefulness of English affixes for predicting lexical categories? We conducted a corpus analysis of English child-directed speech to assess the potential information available in the environment for lexical categorization. A computational system operating optimally is likely to pick up on such signals.

3.1. Method

3.1.1. Corpus preparation

A corpus of child-directed speech was assembled from the CHILDES database (MacWhinney, 2000). We extracted all the speech by adults to children from all the English corpora in the database, resulting in 5,436,855 words. The CHILDES database (with the exception of a minor fragment) provides only orthographic transcriptions of words, so we derived phonological and lexical category information for each word from the CELEX database (Baayen, Pipenbrock, & Gulikers, 1995). Words with alternative pronunciations (e.g., *record*) were assigned the most frequent pronunciation for each orthographic form. Several words had also more than one lexical category. Nelson (1995) showed that in these so-called dual-category words (e.g., *brush*, *kiss*, *bite*, *drink*, *walk*, *hug*, *help*, and *call*) no specific category is learned before the other systematically, but rather the frequency and salience of adult use are the most important factors. Moreover, a well known procedure in computational linguistics that picks the most frequent syntactic category for each word in a corpus is able to tag about 90% of the words correctly (Charniak, Hendrickson, Jacobson, & Perkowitz, 1993). Although 10% of words are still wrongly classified, such a procedure might nonetheless be useful in getting the language system off the ground rather than achieving 100% correct performance. Hence, we assigned each dual-category word the most frequent lexical category from CELEX. For our analysis, we considered the most frequent 4730 words in the CHILDES database.[1] In total, there were 2,541 nouns, 1,108 verbs, and 1,081 other words.

3.1.2. Cue derivation

A comprehensive list of English morphological prefixes and suffixes was compiled, resulting in 248 prefixes and 63 suffixes. Among these, 58 prefixes and 23 suffixes appeared at least once in our corpus. Because some prefixes and suffixes can have more than one phonetic realization (for instance, *-ed* is pronounced /d/ or /t/), we obtained 62 phonetic prefixes and 37 phonetic suffixes. Each word in the corpus was represented as a vector containing (62 + 37) 99 units. If the word started and ended with one of the affixes, then its relevant unit in the

Table 1
Partial vector representation of words in CHILDES based on a subset of two prefixes and two suffixes as predictors

Word	Phonetic Transcription	Lexical Category	Rescrambled Category Assignment (Baseline)	de–	pre–	...	–s	–al
Prepared	/pripærd/	Other	Noun	0	1	...	0	0
Does	/dʌz/	Verb	Other	0	0	...	1	0
Sandal	/sændɬ/	Noun	Verb	0	0	...	0	1
Magical	/mædʒikɬ/	Other	Noun	0	0	...	0	1
Gel	/dʒɛɬ/	Noun	Verb	0	0	...	0	0

vector was assigned a 1, otherwise it was set to 0. At the end of the coding, each word in the corpus consisted of a 99-cue vector with most cues having value 0 and one or two having value of 1. More important, we tested a situation in which the model knows about affixes but knows nothing about their morphological relations to lexical categories. The model simply assigns each word to a lexical category based on its affix. For instance, *-al* as an adjectival suffix will apply both to adjectives like *magical, natural*, and to nouns like *sandal*, *metal*. For the sake of exposition, Table 1 provides a representation of how a sample of words in the corpus would be encoded using a subset of two 2 prefixes and 2 suffixes (in the actual analyses all 99 affixes were represented). In the table, for instance, the words *sandal* and *magical* would be represented by an identical vector "0 0 0 1." An important point is that words that shared a stem (e.g., jam-*s*, jamm-*ing*, jamm-*ed*) were represented as completely separate words. Although it is certainly the case that recognizing that the same stem occurs with different affixes may make the affixes more relevant, such analysis would assume an already sophisticated linguistic partition of words into stem + affix, and the recognition that stems can participate in several stem + affix combinations. Consistently with the idea of a linguistically naïve learner, here we tested the weaker case in which such stem + affix partition may not be known to the child in the earliest phases of syntactic categorization.

To assess the extent to which word prefix and suffix cues resulted in accurate classification, we performed a linear discriminant analysis dividing words into Nouns, Verbs, or Other. Discriminant analyses provide a classification of items into categories based on a set of independent variables. The chosen classification maximizes the correct classification of all members of the predicted groups. In essence, discriminant analysis inserts a hyperplane through the word space, based on the cues that most accurately reflect the actual category distinction. An effective discriminant analysis classifies words in their correct categories, with most words belonging to a given category separated from other words by the hyperplane. We used a "leave-one-out cross-validation" method, which is a conservative measure of classification accuracy, and works by assessing the accuracy of the classification of words that are not used in positioning the hyperplane. This means that the hyperplane is constructed on the basis of the information on all words except one, and then the classification of the omitted word is assessed. This is then repeated for each word, and the overall classification accuracy can then be determined. The results of the analyses of phonological and distributional cues showed that the use of several cues provides not only more accurate classification than

single cues (Monaghan et al., 2005) but also better generalization to novel situations (Reali, Christiansen, & Monaghan, 2003).

Despite its seemingly statistical complexity, discriminant analysis is a simple procedure that can be approximated by simple learning devices such as two-layer "perceptron" neural networks (Murtagh, 1992; see also Experiment 3). Because most cognitive tasks are believed to require at least three-layer neural networks to solve non-linear problems, linear discriminant analyses provide a lower threshold on the type of statistical structure that can be extracted from our word-ending cues.

In line with our suggestion to move toward more plausible assumptions in language modeling, it is reasonable to assume that a young child discovering language will not try to map every new word into fine-grained lexical categories, but will rather start assigning candidate lexical items to broad categories that do not completely correspond to adult lexical categories (Nelson, 1973). In addition, the first adult-like lexical categories will be the most relevant to successful communication. There now exists considerable experimental evidence that children first learn nouns and verbs across languages (Gentner, 1982). The specific number of word classes needed for in a given language is controversial but nouns and verbs seem almost the necessary word classes present in most world languages (Sapir, 1921). Hence, the task of the discriminant analysis was to classify the whole corpus into three categories: Nouns, Verbs, and Other. This classification plausibly reflects the early stages of lexical acquisition, with Other being an amalgamated "super-category" incorporating all lexical items that are not nouns or verbs.[2]

In evaluating the true contribution of morphological cues to classification, one should take into account that a certain percentage of cases could be correctly classified simply by chance. Thus, in order to establish the chance rate a baseline condition was obtained using Monte Carlo simulations. The file containing the data from the corpus had 100 columns: the 99 columns of binary affix predictors (Independent Variables), plus one column that had dummy variable scores of 1, 2, or 3 for the three lexical categories (Dependent Variable). This last column contained 2,541 values of 1 (Noun), 1,108 values of 2 (Verb), and 1,081 values of 3 (Other). We randomly rescrambled the order of the entries in that column while leaving the other 99 columns (the affix predictors) unchanged. Thus, the new random column had exactly the same base rates as the old column but in random order, whereas the first 99 columns were completely unchanged. The rescrambling maintains information available in the vector space, but destroys potential correlations between specific affixes and lexical categories, and thus represents a baseline "control" condition. We created 100 different rescramblings for the Dependent Variable and tested the ability of the 99 affix cues to predict each one of the rescramblings in 100 separate discriminant analyses. In this way, it was possible to test whether in the experimental condition there was coherent phonological consistency within nouns, within verbs, and within other words or whether a three-way classification of words randomly assigned to the three categories would yield similar classification results.

3.2. Results and discussion

When all affix cues were entered simultaneously, 60.7% of cross-validated words were classified correctly, which was highly significant (Function 1 explained 78.5% of the variance,

$\Lambda = .675$, $\chi^2 = 1{,}836.52$, $p < .001$; Function 2 explained 21.5% of the variance, $\Lambda = .912$, $\chi^2 = 429.74$, $p < .001$). Conversely, the 100 discriminant analyses of the baseline yielded a mean correct classification of 35% ($SD = 4.6\%$) and this score was significantly lower than the morphological classification ($p < .01$). More in detail, 76.9% of nouns were classified correctly and significantly better than the baseline condition (39.1%, $SD = 12.5\%$, $p < .001$). For verbs, 54.4% were correctly classified, compared to 30% for the baseline ($SD = 12.6\%$), and the difference was significant ($p < .001$). For Other, 29% of these words were correctly classified using morphological cues, and this was not significantly different from the baseline (30.6%, $SD = 10.2\%$, $t(99) = 1.56$, $p = .12$). Fig. 1 sums up the results. The Other category is harder to classify because it encompasses a very heterogeneous group of words, including open- and closed-class words.

The percentages reported above give an estimate of the "completeness" of the classification procedure (i.e., how many words in a given category are classified correctly). Completeness is calculated by counting the correct number of words classified in a given category (hits) and dividing them by the total number of words correctly belonging to a category (hits + misses). We further measured the "accuracy" of the analyses for each of the three categories, dividing the hits by the total number of words classified in that category (hits + false alarms). Accuracy and completeness are reported in Table 2, including assessments for the baseline condition.

Stepwise analyses were also conducted to assess which cues are most useful in discriminating nouns, verbs, and other classes. At each step, all variables are evaluated to determine which one contributes most to the discrimination between groups. That variable is then included in the model, and the process starts again. Percentage of overall classification obtained with the stepwise method was very similar or identical to the discriminant analyses reported above. Of the 99 phonetic affix predictors entered, 24 were useful in lexical categorization,

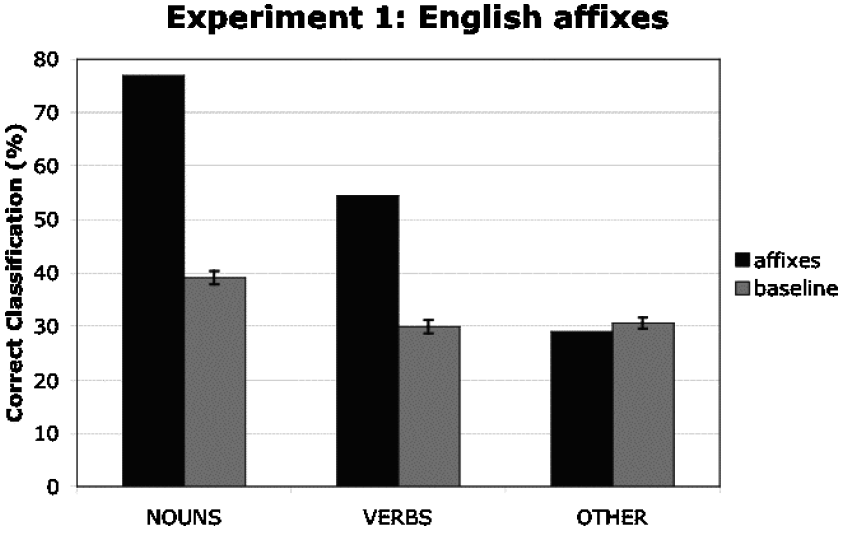

Fig. 1. Percentage of correct classification of English Nouns, Verbs, and Other using morphological information. Baseline classifications are based on 100 Monte Carlo-like simulations. Error bars indicate standard error of the mean. Classification is better than baseline condition for Nouns and Verbs, but not for Other.

Table 2
Summary of accuracy and completeness results for Experiments 1 to 6

	Lexical Category	% Accuracy (Baseline)	% Completeness (Baseline)
English morphological affixes—Experiment 1	NOUN	65 (49)	77 (39)
	VERB	55 (24)	54 (30)
	OTHER	49 (20)	29 (31)
English word-edge phonemes—Experiment 2	NOUN	69 (54)	67 (34)
	VERB	53 (23)	56 (32)
	OTHER	46 (23)	47 (33)
English word edges – Networks—Experiment 3	NOUN	66 (57)	41 (29)
Experimental and baseline groups	VERB	44 (22)	44 (24)
	OTHER	27 (20)	51 (47)
Dutch word-edge phonemes—Experiment 4	NOUN	71 (48)	49 (33)
	VERB	43 (24)	76 (33)
	OTHER	48 (26)	43 (33)
French word-edge phonemes—Experiment 5	NOUN	63 (45)	53 (35)
	VERB	62 (35)	58 (33)
	OTHER	34 (20)	49 (32)
Japanese word-edge phonemes—Experiment 6	NOUN	59 (39)	49 (35)
	VERB	50 (28)	64 (34)
	OTHER	46 (35)	44 (33)

corresponding to the following morphological affixes (in parentheses is the class that they most often predicted, N = Noun; V = Verb; O = Other). The cues are in decreasing order of importance: *-ing* (V), *-ed* (V), *-y* (O), *-er* (N), *-or* (N), *-(o)ry* (N), *-ite* (N), *-id* (V), *-ant* (N), *e-* (N), *-ite* (O), *-ate* (N), *un-* (N), *-ble* (O), *-ive* (O), *an-* (N), *pre-* (N), *out-* (N), *-s (unvoiced;* N), *bi-* (N), *-ine* (N).

The results of Experiment 1 support Maratsos and Chalkley's (1980) intuition that suffixation is a useful cue to lexical category learning in English, and we extended this finding to prefixation as well. Fifteen suffixes and six prefixes were useful in a discriminant analysis (for a discussion of whether prefixes or suffixes are more informative, see section Further analysis II, which follows). It is worth noting that our statistical procedure is in one aspect even simpler than the one envisaged by Maratsos and Chalkley in the first place, because they assumed that the child would perform a type of comparative analysis based on a partitioning of words into "stem + suffix." For instance, upon hearing *walks*, *walked*, and *walking*, the child would analyze "walk + s," "walk + ed," and "walk + ing," with all instances of *walk-* being represented as a single item, the stem "walk." This was not the case in our procedure. Instead, in our vector representation each of the three words above were treated as separate vectors, and no stem + suffix representation was implemented. Therefore, for *walks*, *walked*, and *walking* no part of their vector representations was being shared, corresponding to a representation of "stem = walk." Rather, the representation of the word was "word *walks* ends with -s." Although older children may perform the kind of comparative and analytical analyses proposed by Maratsos and Chalkley, Experiment 1 shows that it is possible to obtain good lexical classifications with

a less analytical procedure that treats each word as a separate entry and attempts to assign it a linguistic label based on its prefix or suffix alone.

However, this procedure has some limitations in terms of plausibility. It assumes that in order to utilize bound morphemes the functional elements themselves must have been previously parsed from the input and represented as linguistic units by the learner. In Experiment 2, we try to overcome such limitation and explore whether an even simpler procedure based on phoneme information can get comparable results to Experiment 1.

4. Experiment 2: A linguistically naïve analysis of word beginnings and endings

Experiment 1 suggests that bound morphemes are potentially useful cues for discovering broad lexical categories. However, the method used implies an already sophisticated level of linguistic analysis where suffixes are pre-analyzed units of the lexicon. Thus, one potential objection to these analyses is that children are not spoon-fed a list of relevant morphological suffixes. We are not discounting the importance of morphology to language acquisition. Indeed, there is some important evidence that the vocabularies of school-age children grow at even a faster rate than preschoolers due to mastery of derivational processes. These allow new words in same and different lexical categories to be derived fast (e.g., do → *un*do; drink → drink*able*; Anglin, 1993; Clark, 1995). Derivational processes also contribute to increasing reading skills (Carlisle & Fleming, 2003).

There is also some evidence for the use of bound morphemes in younger children. A few recent studies have started exploring at which age learners possess bound morphemes as parsed units and whether they use them to assign grammatical categories to the words they modify. Golinkoff, Hirsh-Pasek, and Schweisguth (2001) tested 18- to 21-month-olds using a preferential looking paradigm. They found that, although these children were not yet producing grammatical morphemes in their own speech, they could discriminate a grammatical morpheme that was correctly attached to a word (*-ing* attached to a verb) from one used incorrectly (*-ly* attached to a verb) and could recognize that a nonsense syllable (*-lu*) is not a grammatical morpheme. On the basis of their findings, Golinkoff et al. argued that children may already be distinguishing which morphemes appear with which form classes and that perhaps they show an already sophisticated linguistic analysis of words into stems and morphemes. Mintz (2004) also investigated the learning of the bound morpheme *-ing* in a younger population of 15-month-olds. Using the Head-Turn Preference procedure, he found that infants listened more to nonce pseudo-words (e.g., *gemont*) when they had previously experienced them attached to the morpheme *-ing* (e.g., *gemonting*) than attached to the non-morphemic ending *-dut* (*gemontdut*). Mintz argued that *-ing* contributed to representing the nonce words as stems (i.e., as linguistic units). However, Mintz noted that his preliminary study does not say anything about what kind of status *-ing* has for 15-month-olds (i.e., infants may be sensitive to *-ing* because it is highly frequent in the input and it occurs in many different environments). In addition, both studies mentioned above only deal with the case of the *-ing* morpheme; hence it is unknown at present whether other morphemes are segmented at early stages of language development and assist lexical categorization. Therefore, although

the full morphological system may be developing at later stages and serves an extremely useful function, it may not directly and completely assist the process of learning lexical categories. Does it mean that the beginnings and endings of words are not usable cues?

By 1 year, infants will have learned a great deal about the sound structure of their native language (for reviews, see Jusczyk, 1997, 1999; Kuhl, 1999; Pallier, Christophe, & Mehler, 1997; Werker & Tees, 1999). Thus, when they face the problem of learning lexical categories at the beginning of their second year, they are already well attuned to the phonological regularities of their native language. In particular, infants and children are highly sensitive to word endings (e.g., Slobin, 1973) and word beginnings (Slobin, 1985). Peters (1985) also suggested that learners may use the first and last syllables of larger speech units to start bootstrapping words in the language. The ending phonetic form of words has also been invoked in processes of compound formation (Haskell, MacDonald, & Seidenberg, 2003). Recent experimental work in adult word learning also found a primacy and recency facilitation effect: Adults repeated the beginning and end of nonwords more accurately than the middle of words (Gupta, 2005). Because nonwords are for adults what new words are for children, a reasonable assumption is that whatever sequencing mechanism is responsible for word learning, it displays a learning bias for the beginning and ending of words. There is also evidence that the orthographic beginning and ending segments of words provide useful information that can be integrated in part-of-speech taggers (Mikheev, 1997). We, therefore, developed a simple procedure that children could plausibly use to discover word-edge cues without prior knowledge of morphology and tested its classification success. We chose the phoneme as the unit of analysis because children can distinguish between minimal word pairs involving a single phoneme change as early as 12 months of age (see Jusczyk, 1997). Hence, in evaluating the usefulness of this source of information it would appear that, unlike affixation, it is perceptually available to the child relatively early in development.

4.1. Method

4.1.1. Corpus preparation

The same corpus from Experiment 1 was used.

4.1.2. Cue derivation

We extracted all first and final phonemes from the words in the corpus. By selecting the smallest phonological unit, this procedure makes minimal assumptions about the perceptual and processing capacities of children. Our procedure resulted in 40 beginning and 40 ending phonemes, which combined to form an 80-unit (40 + 40) vector for each word as in Experiment 1. Table 3 shows how the same words as in Table 1 would be assigned to a word-edge vector representation based on only 4 cues (again, this is a fictitious and schematic representation for explanatory purposes—the actual words in the corpus had a 80-bit vector representation). With respect to Experiment 1, most words were assigned a different vector, for instance in Table 3 *does* is represented as "1 0 1 0" using word beginning and ending phonemes, whereas in Table 1 it was represented as "0 0 1 0" using bound morpheme information. This changed the representational space of the corpus. The vectors were entered in a discriminant analysis where the cues were the independent variables and classification for Nouns, Verbs, and Other

Table 3
Vector representations for five words in the corpus, based on a subset of word-edge used in the linguistically naïve analysis of Experiment 2

Word	Phonetic Transcription	Lexical Category	Rescrambled Category Assignment (Baseline)	/d/-	/p/-	...	–/z/	–/ɬ/
prepared	/pripærd/	Other	Noun	0	1	...	0	0
does	/dʌz/	Verb	Other	1	0	...	1	0
sandal	/sændɬ/	Noun	Verb	0	0	...	0	1
magical	/mædʒikɬ/	Other	Noun	0	0	...	0	1
gel	/dʒɛɬ/	Noun	Verb	0	0	...	0	1

Note. Although in Experiment 1 (see Table 1) cues are word prefixes and suffixes, in Experiment 2 they are word-edge phonemes. As a consequence, the same words in the two experiments may have quite different vector representations.

was estimated as in Experiment 1. The baseline condition was estimated using the same Monte Carlo-like procedure as in Experiment 1.

4.2. Results and discussion

An overall 59.7% of cross-validated words were classified correctly, which was highly significant (Function 1 explained 69.7% of the variance, $\Lambda = .667$, $\chi^2 = 1{,}911.35$, $p < .001$; Function 2 explained 30.3% of the variance, $\Lambda = .879$, $\chi^2 = 609.79$, $p < .001$), and was higher than the 100 baseline analyses, which yielded a mean overall classification of 33.7% ($SD = 1.5\%$). More specifically, 66.8% of nouns were correctly classified versus a baseline of 34.4% ($SD = 3.8\%$). For verbs, 56.0% were correctly classified versus a baseline of 32.3% ($SD = 4.2\%$). Last, 47.1% of other words were correctly classified versus a baseline of 33.4% ($SD = 4.2\%$). Student t tests between word-edge and baseline classifications were significant for the overall classification ($p < .01$); and highly significant for nouns, verbs, and other ($p < .001$). The results are summed up in Fig. 2. Accuracy and completeness are reported in Table 2. In stepwise discriminant analyses 26 (10 beginnings and 16 endings) out of the 80 word-edge cues were relevant for successful lexical categorization (Tables 4 and 5).

The results reported here are based on word types (i.e., the frequency of each word does not impact classification) rather than implicitly by the fact that the words are chosen among the most frequent ones in the parental output. However, the same discriminant analyses weighted for the log frequency of each word were not significantly different from the type analyses, $\chi^2(1, N = 9460) = 0.27$, $p = .6$.[3] Experiment 2 suggests that it is possible to achieve good lexical classification in English based on simple word-edge information, namely the beginning and ending phonemes of a word. More important, classification based on morphological cues (Experiment 1) was *not* significantly different from the word-edge classification, $\chi^2(1, N = 9460) = 1.02$, $p < 1$, despite the fact that the morphological classification has 19 additional dimensions among which to carve up word space. Hence, the simpler word-edge model does not loose predictive power. Although it predicts fewer correct nouns than the morphological model, it gains in predicting words in the Other class significantly better than the morphological

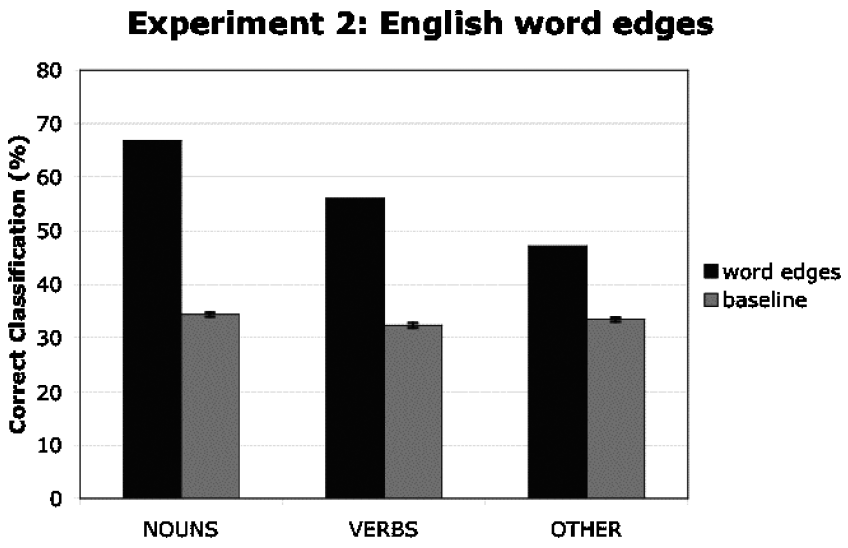

Fig. 2. Percentage of correct classification of English Nouns, Verbs, and Other using first and last phoneme information. Baseline classifications based on 100 Monte Carlo-like simulations (error bars are standard error of the mean).

model. It has to be noted that there is considerable overlap between the word-edge cues and the morphological cues entered in the stepwise analysis of Experiment 1. If only the first phoneme of the prefixes and the last phoneme of the suffixes entered in Experiment 1 was taken into account, 9 phonemes (2 beginnings and 7 endings) from Experiment 1 would overlap with

Table 4
Word-edge cues entered in the stepwise discriminant analysis

IPA Phoneme	Example Word	IPA Phoneme	Example Word
-/s/	Else, this	-/g/	Frog, big, egg
-/ŋ/	Walking, ending	/ə/-	About, another
-/d/	Would, read	-/t/	About, don't, right
-/z/	Does, please, his	/aɪ/-	I
/ɛ/-	Any, egg	-/b/	Scrub, tub
-/tʃ/	Watch, much, which	/n/-	Nanny, know, not, now
/m/-	Mummy, more, make	/d/-	Dear, don't, drink
-/n/	One, can, then	-/ʃ/	Wash, fish, push
-/u/	Shoe, into, blue, through	/v/-	Visit, very, village, visit
-/ɬ/	Animal, purple, trouble, little	/s/-	Same, so, stop, school, see
/p/-	Park, pretty, pick	-/f/	If, enough, yourself, cough
/w/-	Why, one, with, way	-/ɔ/	Draw, saw, law, straw
-/**w**/	Show, how, know	-/ð/	With, clothe, breathe

Note. A hyphen preceding a phoneme signals a word beginning, whereas one that follows a phoneme signals a word ending. Cues are in decreasing order of importance. Pronunciations, derived from CELEX, are based on standard British English. IPA: International Phonetic Alphabet.

Table 5
Number of words in the corpus predicted in each lexical category

Beginning Phoneme	Predicts Noun	Predicts Verb	Predicts Other	Ending Phoneme	Predicts Noun	Predicts Verb	Predicts Other
/ə/	0	4	**51**	-/ɔ/	**8**	0	1
/aɪ/-	**8**	2	2	-/ð/	**3**	0	1
/d/-	**168**	71	16	-/tʃ/	**42**	0	8
/ɛ/-	1	2	**52**	-/ŋ/	7	**368**	0
/m/-	**153**	55	23	-/ɬ/	**101**	0	25
/n/-	7	13	**108**	-/ʃ/	**37**	0	6
/p/-	**289**	43	26	-/b/	**24**	0	2
/s/-	**364**	180	48	-/d/	8	**329**	24
/v/-	**31**	1	3	-/f/	**30**	0	15
/w/-	14	70	**131**	-/g/	**26**	0	5
				-/n/	**217**	0	56
				-/s/	**310**	0	67
				-/t/	69	**379**	59
				-/u/	5	**38**	10
				-/z/	**525**	0	139

Note. Numbers in bold indicate the dominant category.

the word edge phonemes of Experiment 2: -/ŋ/, -/ɬ/, -/d/, -/n/, -/s/, -/t/,-/z/, /ə/-, /p/-. The fact that more than a third of the phoneme cues useful in Experiment 2 are also part of English affixes suggests that affixes contain partial phonological information in simplified form that could be exploited *before* the child acquires knowledge of the morphology of her language. This process of simplification may be akin to visual development in the child, in which the limited acuity in the first 6 months of life acts as a processing filter to help learn about the most relevant close environment first (French, Mermillod, Quinn, Chauvin, & Mareschal, 2002).

An interesting aspect of the word-edge cues is their surprising informativeness despite their potential ambiguity. Table 4 reports word-edge cues relevant to correct classification with example words. From the examples, it is apparent that all cues are potentially highly ambiguous, in that no phoneme clearly signals the beginning or ending of only one lexical category. Table 5 shows in more detail how many Nouns, Verbs, and Other were classified using a specific beginning and ending cue. In the table the informativeness of the cue seems to be given by its classification of words into predominantly—although not exclusively—one of the three classes.

Our discussion so far has revolved around the usefulness of single cues. However, for the purposes of classification, it is important to remember that beginnings and endings may act in concert. Hence, it is worth looking at the word-edge frames that most contributed to correct classification. The theoretical space of possible word-edge frames is large: With 40 beginnings and 40 endings there are $(40 \times 40) = 1{,}600$ possible word-edge combinations (i.e., English words could start with and end in 1,600 possible different ways). Of these, only 132 unique frames were attested in our corpus. Specifically, considering the 26 cues entered in the stepwise analysis, there were a total of 101 relevant "word-edge frames." The Table 6 reports

the 25 most frequent ones (frequency >10), together with the category that they predicted and the distribution of words in the corpus that have the frame and their category in the corpus. The table shows that in most cases the frames categorize according to the most frequent category. Thus, all the 56 verbs starting with /s/ and ending in /ŋ/ were classified correctly. The frames are by no means associated to one specific category: 10 nouns, and 2 other words have an /s–ŋ/ frame, but they were misclassified as verbs. The classification thus reflects a main tendency of frames to be associated primarily with one lexical category. Only four frames go against this trend. /s–t/ and /d–t/ were used in the discriminant analysis to classify verbs, although in

Table 6
Frequent word frames and their predictive power in assigning lexical category

Most Frequent Word-Edge Frame	Category Predicted by Discriminant Analysis	Number of Words in the Corpus That Have Frame and Their Category		
		Noun	Verb	Other
/s–ŋ/	V	10	**56**	2
/s–z/	N	**49**	15	9
/p–z/	N	**45**	9	2
/s–s/	N	**36**	5	5
/p–s/	N	**35**	4	2
/p–t/	N	**30**	12	6
/s–t/	V	26	**28**	10
/d–z/	N*	0	3	**13**
/s–d/	V	9	**24**	6
/m–z/	N	**23**	3	10
/p–ŋ/	V	3	**20**	0
/s–n/	O	4	4	**19**
/w–ŋ/	V	3	**19**	0
/n–z/	O*	**18**	2	6
/d–ŋ/	V	2	**17**	3
/p–n/	N	**17**	0	1
/m–n/	N	**15**	0	1
/m–s/	N	**15**	0	2
/d–t/	V*	**12**	11	5
/d–n/	N	**12**	4	3
/m–ŋ/	V	3	**11**	2
/p–d/	V	5	**11**	3
/p–ɬ/	N	**11**	1	2
/d–d/	V	2	**10**	3
/d–s/	N	**11**	10	3
/m–t/	V	7	**10**	3
/w–z/	O*	**20**	5	10

Note. Most frequent phonetic word-edge frames obtained by the most relevant phonemes entered in the discriminant analysis of Experiment 2. Asterisks on categories in Column 2 individuate when the predicted lexical category is not the most frequent given a specific word-edge frame. Numbers in bold indicate the dominant category in the corpus.

the corpus a roughly equal number of nouns and verbs use these frames. For /n–z/ and /w–z/ the classification reflects one of the two less frequent categories.

Experiments 1 and 2 have established the potential usefulness of word beginnings and endings in supporting lexical categories in English. In particular, Experiment 2 established that a linguistically naïve learner with no prior knowledge of morphological structure may start discovering English lexical categories based on word edge information. This is particularly striking, given that several sounds are ambiguous (/s/ in English may indicate a third person singular on a verb or the plural of countable nouns), and that several sounds entered as cues do not carry any specific morphological meaning (e.g., beginning /h/ was the 11th cue entered in order of importance in the stepwise analysis, although it does not correspond to any morphological prefix in English). Discriminant analysis, however, is a supervised method requiring that the categories are present in the input, and does not allow us to test whether word-edge cues can be useful in generalizing to unlabelled words. For this purpose, we present a largely unsupervised connectionist model in Experiment 3.

5. Experiment 3: Using word edges to predict lexical categories for unseen words

The analyses so far have provided a way to estimate the usefulness of both morphological affixes and word-edge segments, but they do not represent an explicit model for lexical category bootstrapping. This is because in the "leave-one-out" cross-validation method of the discriminant analysis the category information of all but one word is supplied to the model, along with the cues in question, and the model categorizes the left-out word based only on the supplied cues and the discriminant function derived from the words whose categories were known. This process is carried out iteratively so that only one word is being left out at each iteration. In other words, for any given word whose category was unknown, the category of all the other words was known. In this respect, knowledge of the categories must be derived from other sources, and the word edges can at best only support the categories once they are in place. In an attempt to show how word edges can be informative in supporting the identification of lexical categories for unseen words that are syntactically unlabelled, we conducted a series of connectionist simulations in which networks were trained on only a small proportion of the corpus and tested on predicting the lexical category for the majority of the remaining corpus. The rationale is that the small training subset was tagged with lexical information, on the assumption that children may have access to lexical categories for a very limited vocabulary of frequent words. Knowledge of our tripartite distinction between nouns, verbs, and other could be reasonably based on semantic cues, such as the mapping between objects with nouns and events and actions with verbs (Pinker, 1984), and the rest of the lexicon grouped as temporarily belonging neither to objects nor to events or actions. Beside semantic cues, social cues (Tomasello, 2000) might scaffold lexical knowledge of an initial lexicon. A set of simple one-layer perceptrons was trained given this initial state. The results show that using only word-edge segments the networks could predict—with a reasonable degree of accuracy—the lexical category of a large corpus that was unseen during training and *unlabelled* with lexical category information. Therefore, the connectionist simulations demonstrate how it is

possible to utilize word-edge information to generalize to unseen words. In this respect, they provide an early model of lexical category acquisition, complementing the results offered by the discriminant analyses.

5.1. Method

5.1.1. Corpus preparation

The same corpus from Experiments 1 and 2 was used.

5.1.2. Cue derivation

The same 80-unit vector representation for word edges was used as in Experiment 2.

5.1.3. Network architecture

Single-layer perceptrons with 80 input units encoding the 80 word edge cues and 3 output units encoding Noun, Verb, and Other category were trained using steepest descent and a learning rate of 0.1. Each network was initialized with random weights between –0.1 and 0.1. During training, weights were updated after each word presentation but kept frozen during testing. The error derivatives were scaled by the log-frequency of each word in the training set, thus taking into consideration the relative frequency of the words in the corpus (Plaut, McClelland, Seidenberg, & Patterson, 1996). The use of log frequency is common in connectionist modeling (e.g., Harm & Seidenberg, 1999; Seidenberg & McClelland, 1989), and allows learning to be sensitive to token frequency information while preventing low-frequency tokens from being swamped by high-frequency items.

5.1.4. Procedure

Three groups of networks were trained: Experimental, Baseline, and Shuffled. All networks in the three groups were trained on 500 word types and tested on the remaining 4,230 words in the corpus. The training and testing sets were different for the three groups: In the Experimental group, the 143 most frequent nouns, the 142 most frequent verbs, and the 215 most frequent other words were presented to the network. These words formed the top-500 most frequent words in the corpus. The remaining words in the corpus were presented at test. The Baseline group is equivalent to the randomization procedure used in the discriminant analyses. The training and test sets were identical to the Experimental group, but the 500 output vectors from the training set encoding the lexical category information were reassigned randomly to the training items, and the same was done for the 4,230 words in the test set. Thus, the Baseline group provided a way to estimate the learning of lexical categories when their association with the word-edge cues is random. In the Shuffled group, 143 nouns and 142 verbs, and 215 words were randomly picked among the 4,730 words to constitute the training set, and the remaining words formed the test set. The Shuffled group controlled for whether any learning result in the Experimental group was due to the specific top-500 words being used or whether, in fact, any shuffling of 500 words would produce comparable results.

Training consisted in learning the mapping between a word's edges and one of three lexical categories. Testing consisted in predicting one of the three categories based on an unseen word's edges after training. The output unit with the highest activation was selected

as the response for each word presentation in the test phase, and completeness and accuracy was calculated for each network at the end of the test. Mean scores were then computed across networks in each of the three groups. In the Experimental group 10 networks were trained with different random initializations of weights. In the Shuffled group there were 10 different reshufflings for the training and test set, and 10 randomly initialized networks for each reshuffling (the same initial weights as in the Experimental group), resulting in 100 (10 $\times$ 10) separate networks. In the Baseline group, there were 10 different random reassignments of the output labels to the input vectors, and for each reassignment 10 different randomly initialized networks were run (with the same initial weights as in the Experimental group), resulting in 100 (10 $\times$ 10) separate networks.

5.2. Results and discussion

The Experimental networks yielded a mean correct classification of 41% for Nouns, 45% for Verbs, and 51% for Other. The Shuffled networks yielded a mean correct classification of 32% for Nouns, 51% for Verbs, and 64% for Other. The Baseline networks yielded a mean correct classification of 29% for Nouns, 24% for Verbs, and 47% for Other.[4] These results are summed in Fig. 3, and Accuracy is reported in Table 2. Student t tests calculated between the Experimental group and the Baseline group resulted in a highly significant overall difference in performance ($p < .001$), a marginally significant difference for Nouns ($p < .06$), a highly significant difference for Verbs ($p < .001$), and a nonsignificant difference for Other ($p = .6$). Therefore networks that used the correlation between word edges and lexical category for a small portion of the corpus were better at classifying novel Nouns and Verbs in the rest of the corpus than a baseline condition in which such correlation was

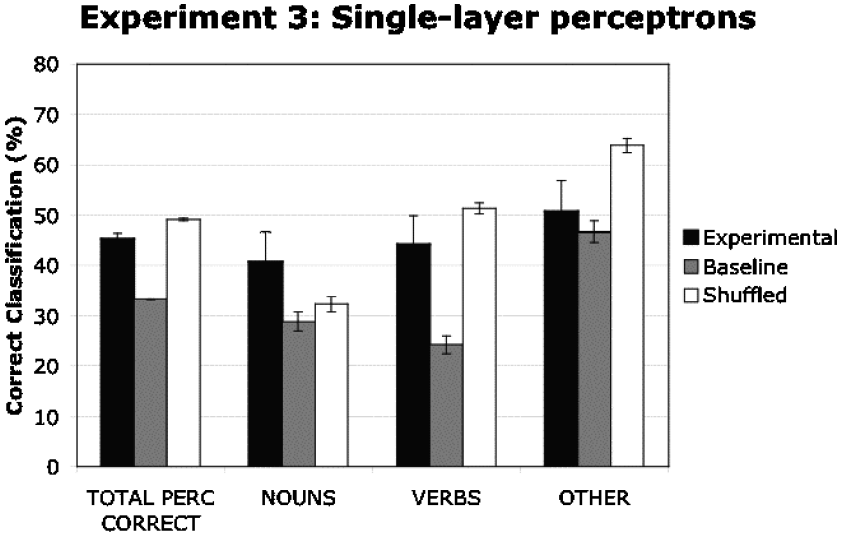

Fig. 3. Percentage of correctly predicted Nouns, Verbs, and Other words in English by single-layer perceptrons. Prediction is based on first and last phoneme information (error bars indicate standard error of the mean).

removed. The results show the potential usefulness of word edges in generalizing to discover lexical categories for novel words that enter the lexicon, and provide an initial model of lexical category acquisition in children. Furthermore, student *t* tests between the Shuffled and Baseline conditions resulted in a highly significant overall difference in performance ($p < .001$), a nonsignificant difference for Nouns ($p = .16$), a highly significant difference for Verbs ($p < .001$), and a highly significant difference for Other ($p < .001$). In addition, student *t* tests between the Experimental and the Shuffled condition revealed a highly significant overall difference ($p < .001$), with the Shuffled group performing better overall; a nonsignificant difference for Nouns ($p = .1$), a marginally significant difference for Verbs ($p = .07$), with the Shuffled group performing better; and a significant difference for Other ($p < .01$).

The current network results go considerably beyond the discriminant analyses. Although in Experiments 1 and 2 all the words but one were labeled for their lexical category at every step, and a statistical model was built by predicting single words iteratively based on word edges, the results of Experiment 3 show that it is possible to generalize beyond an initial small subset of categorized words, and infer the lexical category of completely novel and untagged words. The network analyses also incorporated word frequency information, whereas the discriminant analyses were conducted on word types.[5] Clearly, the results are suboptimal, as not the whole corpus is correctly classified. However, considering the reduced training set, generalization is quite impressive. In addition, we see word-edge cues as but one of several probabilistic cues that in combination can provide a potentially solid scaffolding for language acquisition. Last, we reiterate that the purpose of simple heuristics such as the word edges proposed here is not to induce full-blown adult language knowledge, but instead to get the system "off the ground" initially. We now continue testing our word-edge procedure by determining its generalizability to languages other than English. Because of the similar results between the discriminant analyses and the connectionist model, we use discriminant analyses for the remaining set of experiments.

6. Experiment 4: Dutch

To test the broader applicability of our word-edge procedure to languages other than English, we first extend it to Dutch, a language with structural properties similar to English in several respects. For instance, it is a stress-based language and has a similar morphology. Historically, Dutch and English both descend from the same tree of West Germanic languages, originally spoken by the Germanic speaking people who occupied the southwestern part of the Germanic homeland (Nielsen, 1989).

6.1. Method

6.1.1. Corpus preparation

The Dutch corpus was comprised of the 915,302 word tokens of adult-to-adult and adult-to-child speech from the CHILDES Dutch corpus. The 5000 most frequent words were assigned a phonological representation and a lexical category using the CELEX database.

Words belonging to more than one lexical category were assigned the most frequent category. There were 2,475 nouns, 1,204 verbs, and 1,321 other words.

6.1.2. Cue derivation

Using the same procedure as in Experiments 2 we extracted 37 beginning phonemes and 27 ending phonemes. Each word in the corpus was turned into a 64-unit (37 + 27) vector and entered into a discriminant analysis. The 37 + 27 beginnings and endings were used as predictors in a three-way lexical category classification (Nouns, Verbs, Other). A baseline condition was also established using the Monte Carlo-like procedure of Experiments 1 and 2.

6.1.3. Results and discussion

An overall 54% of cross-validated words were classified correctly (Fig. 4), which was highly significant ($\Lambda = .707$, $\chi^2 = 1{,}725.088$, $p < .001$). This was also significantly better than the baseline condition based on 100 baseline simulations, which yielded a mean correct classification of 33.3% ($SD = 1.8\%$). In particular, 49.3% of nouns (baseline 33.3% , $SD = 4.9\%$), 76.2% of verbs (baseline 33.2%, $SD = 5.2\%$), and 42.6% of other words (33.3%, $SD = 5.5\%$) were correctly classified using the first and last phoneme as word class predictors. Student *t* tests between word-edge and baseline classifications were significant for the overall classification ($p < .01$); and highly significant for nouns, verbs, and other ($p < .001$). Accuracy and completeness are reported in Table 2. Stepwise discriminant analyses revealed that 30 out of the 64 cues (19 beginnings and 11 endings) were relevant for successful lexical categorization (in decreasing order of importance): -ə, -t, -χ, ɛ-, p-, k-, -n, ɔ-, -χ-, Œy-, -r, s-, b-, t-, -ɛi, -f, r-, m-, ə-, ŋ, -s, f-, ʃ-, u-, -i, e:-, n-, ɑ-, -l, j-.

The results of Experiment 3 suggest that the word-edge procedure generalizes well to Dutch. Performance is comparable to English in Experiment 2, although the cues are largely

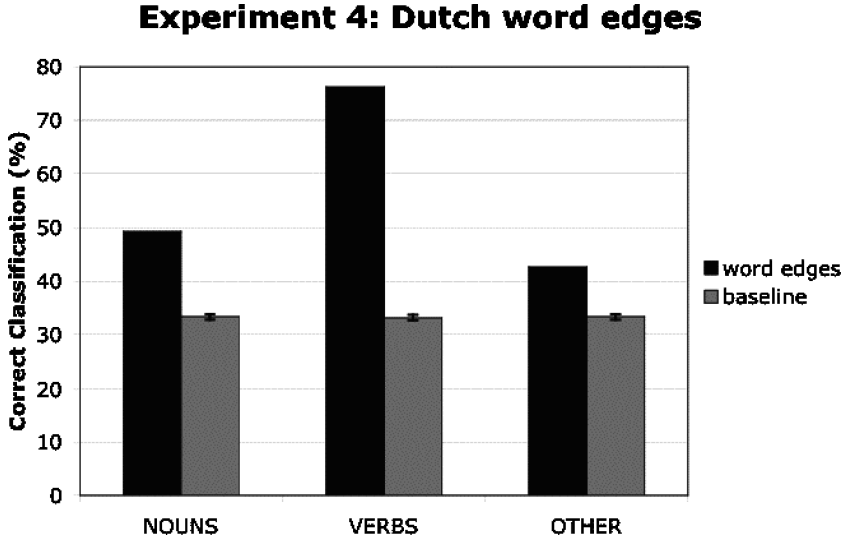

Fig. 4. Percentage of correct classification of Dutch Nouns, Verbs, and Other using first and last phoneme information (error bars indicate standard error of the mean).

different. A notable difference is that beginning cues seem to be more important in Dutch than in English, as a larger proportion of beginnings were entered in the stepwise analysis than endings in Dutch, whereas the opposite occurred in English. Another difference between Dutch and English is that cues worked better for nouns than for verbs in English, but better for verbs than for nouns in Dutch. Although the interpretation of this result is not straightforward, one possibility is that—within a multiple-cue integration perspective—some *other* cues to lexical categorization not considered here may be stronger for verbs in English and for nouns in Dutch. The idea is that if multiple cues to language get integrated during learning, each cue's contribution will have a different weighting to learning depending on the structure, typology, and history of each specific language, but the contribution of the combined constellation of cues is expected to be substantial for learning to take place in any language. In the next Experiment, we extend the word-edge procedure to a non-Germanic language.

7. Experiment 5: French

French has a different historical development than English and Dutch, originating from Latin and belonging to the Romance languages (M. Harris & Vincent, 1990). For our analyses, French is particularly interesting because it is a richly inflected language. Hence, a priori we would anticipate that a great deal of information is available at the edge of French words. At the same time, many word endings signal different categories (e.g., *fait* = noun/verb, *fais* = verb, *mais* = preposition, *lait* = noun, all end with the closed vowel /e/; *allons* = verb, *son* = noun/adjective, *sont* = verb *on* = pronoun all end with the nasal vowel /õ/). In addition, the French lexicon possesses a larger number of homophones than English (Gauvain, Lamel, Adda, & Adda-Decker, 1993) These homophones belong to different lexical categories and cause more errors in French than English speech recognition systems (Gauvain et al., 1993; e.g., *elle* = she = pronoun, is pronounced like *aile* = wing = noun; the sound /ɛ/ stands for the following: *ai* = first person singular present indicative of *avoir* (to have), *aie* = first person singular subjunctive of *avoir; aient* = third person plural subjunctive of *avoir*, *aies* = second person singular subjunctive of *avoir*; *ait* = third person singular subjunctive of *avoir*; *es* = second person singular present indicative of *être* (to be); *est* = third person singular present indicative of *être*; *et* = and, conjunction). French is also a syllable-timed language, and may thus potentially contain more information at the syllable level than at the phoneme level. Therefore, it is not clear a priori whether word-edge phonemes may be useful in classifying words in French. We investigated this by running discriminant analyses on a corpus of child-directed French.

7.1. Method

7.1.1. Corpus preparation

Child-directed speech from the French subcorpora of CHILDES was extracted and its 3,000 most frequent words (amounting to 353,260 word tokens) were assigned a phonological representation and a lexical category using the LEXIQUE database (New, Pallier, Ferrand, & Matos, 2001). In case of dual-category words (e.g., *fait* = noun,verb) the most frequent

category was assigned as in previous experiments. There were 1,360 nouns, 1,053 verbs, and 587 other words.

7.1.2. Cue derivation

The same procedure extracting the first and last phoneme of each word in the corpus was adopted, resulting in 37 beginnings and 36 endings. Each word was transformed into a 73-unit (37 + 36) vector, and entered in a discriminant analysis where the 73 cues were used as predictors of a three-way lexical category classification (Nouns, Verbs, Other). As in previous experiments, we compared our analyses to a baseline condition using the same Monte Carlo procedure.

7.2. Results and discussion

An overall 53.9% of cross-validated words were classified correctly (Fig. 5), which was highly significant ($\Lambda = .680$, $\chi^2 = 1{,}142.593$, $p < .001$), whereas the overall baseline classification was at 33.8% ($SD = 1.6\%$). In particular, 52.6% of nouns (baseline 34.9%, $SD = 4.1\%$), 57.8% of verbs (baseline 33.2%, $SD = 4.5\%$), and 48.7% of other words (baseline 32.2%, $SD = 4.7\%$) were correctly classified using the first and last phoneme as word class predictors. Lower baseline classifications were significant for overall classification ($p < .01$), and highly significant for nouns, verbs and other ($p < .001$). Accuracy and completeness are reported in Table 2. Stepwise discriminant analyses revealed that 33 of the 73 cues (12 beginnings, 21 endings) were relevant for successful lexical categorization (in decreasing order of importance): -e, -ɛ, R-, -a, a-, ã-, -o, -y, -v, -s, -õ, d-, -ə, h-, o-, n, -Ø, u-, -d, -l, -ɛ̃, v-, s-, ɛ-, -O, -l, -œ, -ʃ, -ɲ, f-, ʃ-, -k.

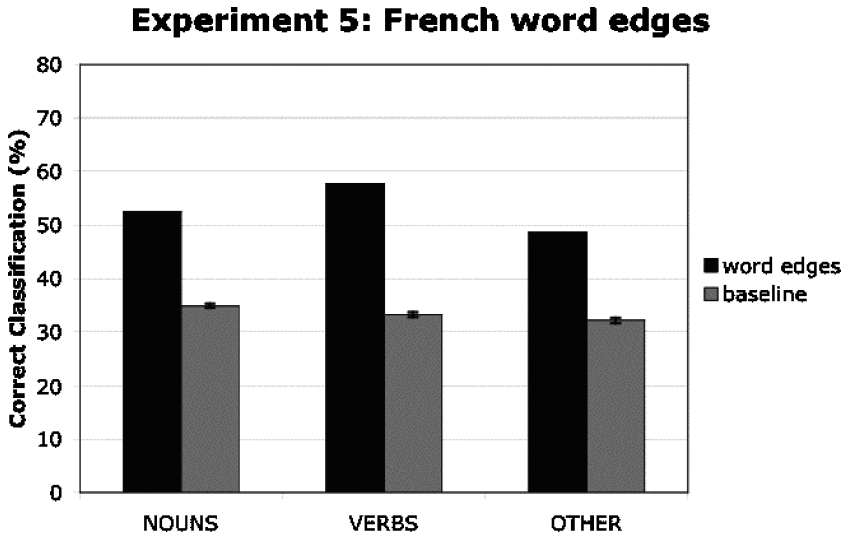

Fig. 5. Percentage of correct classification of French Nouns, Verbs, and Other using first and last phoneme information (error bars indicate standard error of the means).

The results suggest that lexical classification using word-edge French cues is comparable to results in English and Dutch. Hence, the word-edge procedure generalizes well to a non-Germanic language that displays considerable sound ambiguity. Similar to English, the French stepwise analysis involved almost twice as many endings as beginnings, suggesting a greater impact of endings to lexical classification (for more details on this, see Further analyses II, which follows). Like Dutch, classification in French was better for verbs than for nouns. We now turn to the last of the languages investigated in this article, namely Japanese.

8. Experiment 6: Japanese

Our last extension of the word-edge procedure applies to a non Indo-European language. Japanese is an agglutinative language in which bound morphemes are used. For verbs, morphemes are obligatory, in that a verb root is never encountered by itself (Shibatani, 1990) Many bound morphemes in Japanese, in addition, determine the social relationship, age, sex, and status of the speaker and listener, as well as any third parties being discussed and so are heavily present on nouns. Several bound morphemes are also used in a complex system of honorifics, which indicate the relative status of the speaker to the listener, as well as respect (or lack thereof) to the person being spoken of. Therefore, it appears that a priori word beginnings and endings might play an important role in signaling lexical categories (Shibatani, 1990). For the purpose of testing the robustness of our simple word-edge method, however, Japanese represents an interesting case in that only a very limited number of phonemes end words in Japanese, suggesting that most words across categories share the same endings. In terms of rhythmic distinctions, Japanese is a mora-timed language, and this may actually be advantageous for classification as some word-ending morae may coincide with single phonemes. Given these considerations, we conducted a series of discriminant analyses on child-directed speech to test the usefulness of our word-edge learning procedure in Japanese.

8.1. Method

8.1.1. Corpus preparation

Child-directed and adult-directed speech from the Japanese subcorpus of CHILDES was extracted, amounting to 358,401 word tokens. The 1,000 most frequent words were assigned a phonological representation and a lexical category using the CALLHOME corpus (Canavan & Zipperlen, 1996), with hand-assignment for the most frequent 1,000 words by a native Japanese speaker. Words belonging to more than one lexical category were assigned the most frequent category. There were 382 nouns, 276 verbs, and 342 other words. It must be noted that Japanese does not have a standard segmentation strategy, and when adults are asked to segment speech, they do so in idiosyncratic ways, unlike English, French, Dutch adults. The Japanese corpora in Childes (Hamasaki, 2002; Ishii, 1999; Oshima-Takane, MacWhinney, Sirai, Miyata, & Naka, 1998) are parsed according to the Wakachi98 and Wakachi02 procedure (Miyata & Naka, 1998), hence the results reported later apply to this specific procedure.

8.1.2. Cue derivation

The same procedure as in Experiments 2 through 4 was used to extract 29 beginning phonemes and 9 ending phonemes. Each word in the corpus was turned into a 38-unit (29 + 9) vector and entered into a discriminant analysis. As in the previous experiments, the 38 beginnings and endings were used as predictors in a three-way lexical category classification (Nouns, Verbs, Other). A baseline condition was established using the same Monte Carlo baseline procedure as in previous experiments.

8.2. Results and discussion

An overall 51.5% of cross-validated words were classified correctly (Fig. 6), which was highly significant (Function 1 explained 73.4% of the variance, $\Lambda = .703$, $\chi^2 = 345.82$, $p < .001$; Function 2 explained 26.6% of the variance, $\Lambda = .905$, $\chi^2 = 97.32$, $p < .001$), and was significantly higher than the baseline classifications (34%, $SD = 2.5\%$; $p < .01$). In detail, 49% of nouns (baseline 35.3%, $SD = 5.8\%$), 64.1% of verbs (baseline 33.7%, $SD = 6.1\%$), and 44.2% of other words (baseline 32.7%, $SD = 6\%$) were correctly cross-classified using the first and last phoneme as word class predictors. Student t tests between word-edge and baseline classifications were highly significant for nouns, verbs and other ($p < .001$). Accuracy and completeness are reported in Table 2. A stepwise analysis yielded 14 out of 38 relevant word-edge cues (in decreasing order of relevance): -w, ŋ, ɔ-, -ɛ, i-, ʃ, b-, -a, ç-, t-, j-, ŋ-, m-, ϕ-.

The results suggest that Japanese word-edge cues are comparable in performance to cues for English, Dutch, and French (see Fig. 7 for an overall comparison). Like Dutch and French, Japanese cues seem to be more informative in classifying verbs than nouns. In addition, similar to Dutch beginning cues contributed more than endings (10 beginnings and 4 endings were

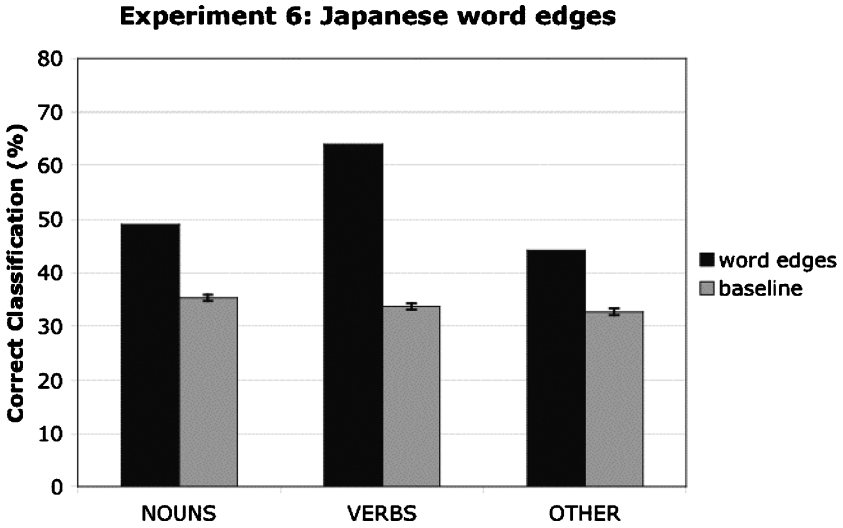

Fig. 6. Lexical categorization of Japanese based on word-edge information (error bars indicate standard error of the mean).

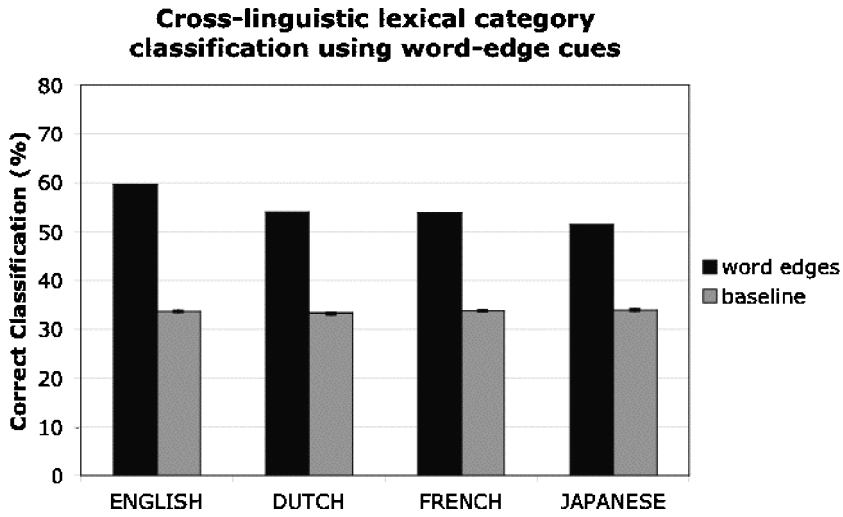

Fig. 7. Comparison of overall percentage of correct classification of Nouns, Verbs, and Other using word-edge information across the four languages studies: English, Dutch, French, and Japanese (error bars indicate standard error of the mean).

entered in the stepwise analysis), an opposite trend to English and French. A notable fact is that it was possible to obtain comparable performance levels in Japanese with a restricted number of cues (38), in fact half the number of cues used in the English word-edge analyses of Experiment 2. The other languages also involved larger vector representations (Dutch, 64; French, 73). Overall, then, the results from the four Experiments suggest that a simple procedure that is only sensitive to phonemes at the edge of words can be as informative as a more sophisticated morphological analysis of the input.

9. General discussion

In language acquisition research, a hypothesis is gaining ground that children may exploit various sources of low-level information available in the input to start individuating structural linguistic relations such as lexical categories. Because most sources of information are probabilistic, it is further hypothesized that the child must ultimately integrate them using simple learning mechanisms. Although the potential importance of word beginnings and endings has long been noted (Maratsos & Chalkley, 1980; Peters, 1985; Slobin, 1973, 1985;), no empirical study has assessed their usefulness in learning syntactic categories, and we decided to make a quantitative estimate based on corpora of child-directed speech of English, French, Dutch, and Japanese. In this article, we have demonstrated that there are effective correlations between word-edge phonetic cues and lexical categories, which might potentially support the development of lexical knowledge.

We have also made suggestions for moving toward more plausible computational analyses of the early stages of language acquisition. Plausibility criteria were introduced from a

developmental perspective. First, language knowledge is constructed progressively and some linguistic categorizations are learned before others. We considered that an initial state of lexical categorization may not involve fine-grained adult-like category distinctions. Rather, lexical categorization may start by distinguishing those categories that children seem to learn first, namely nouns and verbs, perhaps because these categories can map more directly to clear semantic properties of the world, such as objects for nouns and states/actions for verbs (Pinker, 1984). For this reason our analyses involved a coarse distinction between nouns, verbs, and other words, where "other" was a category on its own that would be split into finer-grained categories such as adjectives, determiners, etc. at a later stage. Another dimension of psychological plausibility pertains to the type of input representation considered: In this respect, we attempted to make minimal assumptions about the linguistic units available to the child and their perceptual accessibility at an early stage of development. After assessing the usefulness of linguistically-defined morphological affixes in Experiment 1, we showed that similarly good categorization results can be obtained by a naïve learner that simply focused on the first and last phoneme of a word—a more usable source of information because by their second year of life infants have developed a striking sensitivity to the sound patterns of their language as well as a sensitivity to word beginnings and endings.

Another crucial criterion of plausibility was generalizability from an original subset of learned word-form-to-category mappings to the rest of the lexicon, as well as generalizability of word-edge cues across different languages. Because discriminant analysis is a supervised method that requires providing a lexical category in all but one word at each step we presented a largely unsupervised model in which the size of the pre-labeled corpus was minimal. In Experiment 3 a simple two-layer perceptron generalized its knowledge of word edges to predict the lexical category of unseen words, after being trained on word-to-category mappings for only a small subset of the whole corpus. In other words, the system only needs a limited number of labeled cases to "get off the ground." The source of information for the labeled cases may come from consistent semantic word-to-world mappings (Pinker, 1984), from distributional information (e.g., Redington et al., 1998), or social cues (Tomasello, 2000), or a combination of these cues. Finally, Experiments 4 through 6 proposed that the word-edge procedure extends to three languages other than English, and laid out the basis for future empirical studies with other languages.

Some considerations are in order with respect to word-edge information: In our analyses, we chose the phoneme as the relevant unit. There is evidence that speakers of "stress-timed" languages such as English and Dutch show greater access to phonemes (e.g., Cutler, Mehler, Norris, & Segui, 1986; Cutler & Norris, 1988; Vroomen, van Zon, & de Gelder, 1996). It may be that children are sensitive to other word beginning and ending units larger than the phoneme. For instance, speakers of "syllable-timed" languages (e.g., French, Italian, Spanish, Catalan, and Portuguese) show a processing advantage for syllables (e.g., Mehler, Dommergues, Frauenfelder, & Segui, 1981; Morais, Content, Cary, Mehler, & Segui, 1989; Sebastián-Gallés, Dupoux, Segui, & Mehler, 1992), and Japanese adults use morae as the primary unit of segmentation (Otake, Hatano, Cutler, & Mehler, 1993). Future additional analyses done on such units might shed more light on this aspect. In Experiment 6 it was noted that to the extent that the mora in Japanese may correspond to a single phoneme our analysis is partly coextensive with a mora-unit analysis.

There are other reasons to believe that the phoneme as a unit has a fundamental role in language acquisition. In corpus analyses of English child-directed speech from CHILDES and other corpora Hockema (2006) showed that the speech stream is primarily characterized by phoneme transitions that tend to be of just two kinds: those that occur within a word and those that occur between words. Preliminary results (Christiansen, Hockema, & Onnis, 2006; Christiansen, Onnis, & Hockema, 2008) suggest that a statistical learner that tracked transitional probability information between phonemes would be able to discover the boundaries of lexical units in continuous speech and—because these boundaries largely coincide with the analyses reported in this article—use such boundaries to start determining the lexical category of such units. Hence, statistical information about the distribution of phonemes in the speech signal can contribute to attenuate two major language acquisition problems at the same time: speech segmentation and syntactic category assignment. These analyses thus indicate that the simplification in the current analyses involving a perfectly segmented speech corpus is not necessary for word-edge cues to be useful. Christiansen et al. (2008) outlined a plausible segmentation scenario in which the segmentation outcome was suboptimal but, more important, still provided a solid basis for the discovery of lexical categories using word-edge information. Further delineation of potential developmental trajectories of early word segmentation and lexical discovery is outside the scope of this article, although we are pursuing it in other work. Below, we expand on the results obtained in the discriminant analyses experiments and specify further post-hoc analyses.

9.1. Further analyses I: From usefulness to usability

In discovering language structure from speech the search space is vast. Because specific mappings for a specific human language must be learned by the newborn child, a first step toward reducing the search space is to assess what statistical properties of the specific language are potentially useful and usable. Several studies reviewed here have indicated that the sound properties of words are both useful and used for grammatical categorization. In this article, we have proposed a distinction between *usefulness* and *usability* in computational models of language acquisition. Although many useful sources of information may be present in the raw input from birth, they may not become *usable* until later stages of perceptual development. Therefore, we introduce the notion of *usability* as a series of prerequisites for the psychological reality of language acquisition models. A first distinction between usefulness and usability incorporates a notable distinction between *input* and *uptake*. M. Harris (1992) defined uptake as "that part of the input that is actually attended to by the child." To give an example, based on Santelmann and Jusczyk (1998) the uptake from the input for tracking non-adjacent relations such as *is . . . -ing* at 18 months is limited to three successive syllables (see also Gallaway & Richards, 1994, for a further distinction between uptake and intake).

Several studies have proposed that processing restrictions are actually beneficial to the child, in that they allow focusing on certain basic properties of language upon which to build further language at successive stages (see the "less is more" hypothesis; e.g., Elman, 1993; Newport, 1990). Hence, in our Experiment 2 we showed that cues that are more likely to be used by young children—phonemes at the edge of the word as opposed to full-fledge morphological units—are comparable to morphological information. In addition, the finding

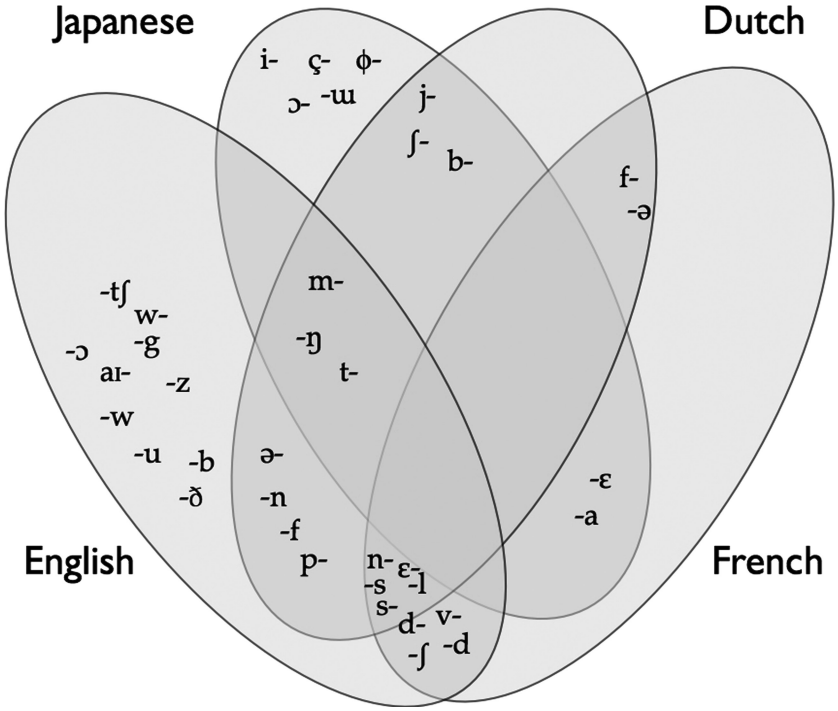

Fig. 8. Venn diagram showing the partial overlap of word-edge cues across the four languages studied.

that word edges are useful in a variety of languages other than English (Experiments 4–6) lends considerable additional credence to the usability of these cues. Specific cues need not be the same across languages: A Venn diagram (Fig. 8) shows the partial overlap of the cues entered in the stepwise analyses for the various languages. More important, no single cue is universally important for all four of these languages.

In pursuing our goal of specifying usability in computational models of language, we were interested in comparing the validity of our word-edge cue procedure with other corpus-based estimates of useful phonological information. In particular, Monaghan et al. (2005) compiled extensive corpus measures of phonological information available in child-directed speech. They used 16 phonological cues that that have been suggested to be relevant for discriminating between noun and verbs and between function words and content words. The cues were at the word level (phoneme length, syllable length, presence of stress, and stress position), at the syllable level (onset complexity, syllabic complexity, reduced syllables, reduced first vowel, *-ed* inflection), and at the phoneme level (presence of coronals, initial unvoiced dental fricative, final voicing, nasal, stressed vowel position, vowel position, vowel height). Monaghan et al. reported classification results of 58.5% for Nouns and 68.3% for Verbs (61.3% overall classification). In order to compare the usefulness of our word edges directly with Monaghan et al.'s results we ran a linear discriminant analysis on the simpler 2-way (Noun/Verb) classification task. Hence, we entered only nouns and verbs in the analysis, leaving out all other words in the corpus. This resulted in 76.9% overall correct classification

(Nouns = 83.5%, Verbs = 61.8%), which was significantly better than Monaghan et al.'s phonological cues ($\chi^2 = 192.12$, $p \leq .001$). Word-edge information requires combination of only two very salient features, the first and the last phoneme of a word, and is arguably a simpler source of information than the 16 combined phonological cues. This fact, coupled with better classification results of the discriminant analyses, increases the usability of word edge information.

In the Introduction, we also argued that a working definition of usability should take into consideration the partial and non-adult-like status of children's initial lexical categories (Nelson, 1973; Tomasello, 2000), in particular when it comes to modeling the very early stages of language development. We, therefore, made the simplifying assumption that children would start by classifying the most relevant lexical categories, nouns and verbs, from the beginning, whereas other categories would be lumped together in a "super category," namely Other. Although Monaghan et al. (2005) also made a similar starting assumption, they excluded the Other category altogether from their analysis. This is equivalent to the child being exposed to a corpus of only nouns and verbs, with all other words being completely wiped out from the input. Although not impossible, this simplification assumes a filtering process, requiring the child first to divide the lexicon into two super-categories—Nouns + Verbs on the one side, and Other words on the other. At a later stage the child would filter out the Other words in order to focus on categorizing nouns and verbs first. This step was not modeled by Monaghan et al. Conversely our three-way classification task maintained a plausible simplified non-adult-like categorization of the lexicon (Noun, Verb, Other) without the need to filter out the input from words "irrelevant" to the task. In this respect, our three-way classification task may be more psychologically plausible. For this reason, we also wanted to compare how our word-edge procedure fared with respect to the 16 phonological cues of Monaghan et al. on the more complex three-way classification task. Hence, we ran a discriminant analysis on our corpus[6] using the 16 phonological cues as independent variables: We obtained an overall 46.4% classification (41.2% Nouns, 60.5% Verbs, and 44.3% Other). Word-edge cues of Experiment 2 were significantly better than Monaghan et al.'s cues on the three-category distinction ($\chi^2 = 186.19$, $p \leq .001$).

From our comparisons with Monaghan et al. (2005), we can draw a series of conclusions: First, the usefulness of Monaghan et al.'s phonological cues was confirmed even in the three-way classification, although clearly to a lesser degree than their original two-way classification. This lower performance is not problematic considering that a large portion of the lexicon was excluded in the two-way classification. Second, our simpler word-edge discovery procedure was better than the 16 phonological cues in both two-way and three-way classification tasks. However, the word-edge classification procedure has the advantage of having five times more dimensions available for carving up the word space, and this is likely to have contributed to the differences in classification performance between the two types of cues. Moreover, the phonological cues may be particularly useful for English verb classification (Christiansen & Monaghan, 2006), something that the word-edge analyses did not indicate for English, but only for Dutch, French, and Japanese. Given that there is little overlap between our word-edge cues and the 16 phonological cues used by Monaghan et al., children could potentially use both types of cues for lexical category discovery. Last, although performance on the three-way classification task was obviously lower than on the two-way classification task for

both phonological cues and word edges, the three-way classification is more psychologically plausible because it does not exclude words from the input.

9.2. *Further analyses II: Differential contribution of beginnings and endings*

An important question in evaluating word-edge information is whether beginnings and endings contribute equally to classification or whether one of the two is more informative and whether this is true across languages. To this end we ran further discrimination analyses using only word-edge beginnings and only word-edge endings respectively as predictors of lexical category. We examined which words were correctly classified by these partial analyses, and also explored which words were correctly classified using beginning cues only but which were incorrectly classified using ending cues, and vice versa, the case where ending cues produced a correct classification but beginning cues resulted in incorrect classification. Finally, we also noted those words that were classified incorrectly by both analyses. There were two possibilities for the resulting classifications. It may be that the same words are correctly classified by analyses based on both cue types, or it may be that there is complementarity in the classifications: Those words incorrectly classified by, say, the beginning cues, may be correctly classified by the ending cues. Table 7 presents the results for English, showing the number of words on which the classifications agreed and disagreed. A hierarchical loglinear analysis were used to assess whether there are main effects and interactions between the classifications based on the different cue types and the Noun/Verb/Other category. One-, two-, and three-way log-linear analyses on the data shown in Table 7 were carried out. The one-way analyses refer

Table 7
Correct and incorrect classifications based on beginnings or endings

		Endings		
	Noun	Correct	Incorrect	Total
Beginnings	Correct	915	344	1,259
	Incorrect	920	362	1,282
	Total	1835	706	2,541
		Endings		
	Verb	Correct	Incorrect	Total
Beginnings	Correct	307	196	503
	Incorrect	372	233	605
	Total	679	429	1,108
		Endings		
	Other	Correct	Incorrect	Total
Beginnings	Correct	105	307	412
	Incorrect	198	471	669
	Total	303	778	1,081

to main effects in the table, the two-way analyses refer to interactions between two of the factors, and the three-way analysis tests whether there is a three-way interaction in the table. The one-way effect of Category (Noun, Verb, Other; $\chi^2(24, N = 4730) = 3{,}104.83$, $p < .001$) can be explained by there being more nouns than verbs and other words. The one-way effects of beginning cues, $\chi^2(24, N = 4730) = 3{,}583.89$, $p < .001$, and ending cues, $\chi^2(24, N = 4730) = 2{,}397.65$, $p < .001$ reflected the fact that each classification assigned words to the correct category significantly more than by chance. The two-way effects of Category by beginning cues, $\chi^2(22, N = 4730) = 2{,}756.24$, $p < .001$, and Category by ending cues, $\chi^2(22, N = 4730) = 1{,}570$, $p < .001$, indicate that the classifications were more successful overall for nouns and verbs than for other words, which is reasonable given that the Other category is a heterogeneous super category. The two-way effect of beginning cues by ending cues, $\chi^2(22, N = 4730) = 2{,}049.05$, $p < .001$, was due to ending cues being more effective in classifying words than the beginning cues. However, interpretation of these lower-level interactions must be moderated by the three-way interaction.

The three-way interaction (Category × Beginning × Ending; $\chi^2(20, N = 4730) = 1{,}221.4$, $p < .001$) suggests that the combination of beginning and ending information operates differently for nouns, verbs, and other. The principal differences in the classifications in Table 7 are the number of words that the beginning and ending cues classify wrongly. For nouns, beginning cues misclassify almost twice (1,282) as many words as ending cues (706), so 72% of Nouns incorrectly classified by beginnings were remedied by endings, whereas 49% of Nouns incorrectly classified by endings were remedied by beginnings. A similar trend can be seen for verbs: 62% of Verbs incorrectly classified by beginnings are remedied by endings, whereas 47% of Verbs incorrectly classified by endings are remedied by beginnings. Hence, it appears that whereas both beginnings and endings contribute to correct classification, endings contribute more for nouns and verbs. For the Other category, however, there is a preferential role for beginnings over endings: 30% of Other words misclassified by beginnings were remedied by endings, and 39% of Other misclassified by endings were remedied by beginnings.

How do beginnings and endings operate in Dutch, French, and Japanese? Do they contribute differentially to lexical categorization as in English, where endings seem particularly informative in determining nouns and verb? If this was true, then a case could be made for the preferential role of ending cues across languages, as Slobin (1973) suggested. Therefore, we conducted one-, two-, and three-way hierarchical log-linear analyses for Dutch, French, and Japanese in the same way as for English above.

For Dutch, all main effects and interactions were highly significant ($p < .001$). Looking at the interactions, the classifications were more successful for nouns and verbs over other words. In addition, 62% of incorrect classifications by endings were remedied by beginnings, whereas 38% of incorrect classifications by beginnings were remedied by endings. For verbs, the opposite pattern applies: Only 26% of incorrect classifications by endings were remedied by beginnings, whereas 84% of beginnings were remedied by endings, suggesting that ending cues are very effective on classification of verbs, whereas beginning cues were particularly effective for nouns in Dutch. For Other words, there was no particular preference for beginning or ending cues, although both still contributed to correct about 30% of each other's misclassifications. This pattern is somewhat different from English, where endings appear more effective for nouns *and* verbs, whereas beginnings work better for other.

For French, all main effects and interactions were significant ($p < .001$). Endings overall contributed more to correct classification than beginnings, although to a lesser extent than English and Dutch. Both beginnings and endings contributed equally to correct each other's noun misclassifications (55% and 53% respectively), while endings contributed more for verbs (correcting 51% of misclassifications by beginnings) and other words (correcting 44% of misclassifications by beginnings). Again, this pattern is slightly different from English and Dutch.

For Japanese, all main effects and interactions were significant ($p < .001$). Beginnings were more useful for classifying nouns and other words and endings were more useful verbs: 51% of nouns, 64% of verbs, and 42% of other words misclassified by endings were remedied by beginnings, whereas 39% of nouns, 75% of verbs, and 32% of other words misclassified by beginnings were remedied by endings. This pattern is more similar to Dutch than to English and French, suggesting language-specific patterns of informativeness of the cues.

In pursuing these detailed investigations of word edges it is useful to keep in mind that the analyses do not tell us directly how the learning mechanism works, but they provide us with useful information to infer how a learning mechanism *should* work if it were to capitalize on word-edge information contained in the input. What the results of the cross-linguistic log-linear analyses reveal can be summed up as follows: First, beginnings and endings help each other in reducing misclassifications, and therefore a learning mechanism that capitalizes on this information for individuating lexical categories should be perceptually attuned to *integrating* both types on information simultaneously from an early stage. The integration part is particularly relevant. This supports earlier postulations of "operating principles" (Peters, 1985; Slobin, 1985) that the learning mechanism should pay attention to the boundaries of speech units. Second, there does not seem to be a universal preference in informativeness attached to the endings of words, at least based on the four languages investigated here, and this runs counter to Slobin's (1973) idea that a general cognitive bias for endings should be in place. Slobin's (1973) account was based on cross-linguistic evidence that locative markers in postverbal and postnominal positions (as in Hungarian) tend to be acquired before ones in preverbal and pronominal positions (as in Serbo-Croatian).

Our analyses and others (e.g., Fisher & Tokura, 1996) seem to suggest that a constellation of cues operate differently for different categories across languages, and as long as these cues *complement* each other when integrated, they will provide a solid statistical scaffolding to structure discovery directly available in the speech signal. Our analyses could be used in further behavioral and computational studies assessing the contribution of other sources of information integrated with word-edge information. We conclude that simple computational principles can be quite powerful even in isolation, although a complete account of language acquisition will require a combination of many simple computational principles for the detection and integration of multiple sources of probabilistic information.

Notes

1. The original corpus consisted of the 5,000 most-frequent words, but many of these did not have a phonetic transcription (they were either proper names, names of toys, or misspellings such as *didn*). After partly hand cleaning, we ended up with a clean

corpus of 4,730 words for which an automatic phonetic transcription could be obtained in CELEX.

2. Infants also seem to discriminate function and content words early in their language development (Shi, Werker, & Morgan, 1999); thus, a further distinction in the Other category would have seemed justified. However, other cues than word beginnings and endings have been assessed as useful in this distinction, notably word length and word-internal cues (Monaghan, Chater, & Christiansen, 2005), which may be integrated with the cues proposed here in language acquisition.
3. This result also applies to the discriminant analyses in Experiment 1. In three separated analyses, word log frequency accounted for between 32.7% and 44.1% of the variance in predicting the age of acquisition of a word, nearly 10 times more than the 3.0% to 4.9% range obtained for raw word frequency. Thus, log frequency provides a reasonable approximation of the word token statistics to which a child is likely to be sensitive (Christiansen, Onnis, & Hockema, 2008).
4. Similar results with minor variations were also obtained with networks trained with different learning algorithms and parameterizations, suggesting that the network results are robust.
5. We also obtained similar results in network simulations when frequency was not included, suggesting that, at least in our model, word frequency plays a minor role.
6. Our corpus is in fact the same corpus used by Monaghan, Chater, and Christiansen (2005) with some minimal modifications, such as spelling corrections on some of the words listed.

Acknowledgments

Support comes from Human Frontiers Science Program Grant RGP0177/2001–B to Morten H. Christiansen. We thank Padraic Monaghan for contributing an early corpus, and Dick Darlington and Thomas Farmer for useful comments on the statistical analyses. We also thank the editor and three anonymous reviewers for important insights that improved the original manuscript. The order of the authors is random.

References

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10, Serial No. 238).

Baayen, R. H., Pipenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bowerman, M. (1973). Structural relationships in children's utterances: Syntactic or semantic? In T. Moore (Ed.), *Cognitive development and the acquisition of language*. Cambridge, MA: Harvard University Press.

Brooks, P. J., Braine, M. D., Catalano, L., & Brody, R. E. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32,* 76–95.

Canavan, A., & Zipperlen, G. (1996). *CALLHOME Japanese speech*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Carlisle, J. F. & Fleming, J. (2003). Lexical processing of morphologically complex words in the elementary years. *Scientific Studies of Reading, 7,* 239–253.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language, 30,* 348–369.

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin and Review, 8,* 519–523.

Charniak, E., Hendrickson, C., Jacobson, N., & Perkowitz, M. (1993). Equations for part-of-speech tagging. In *Proceedings of the 11th National Conference on Artificial Intelligence*, (pp. 784–789). Washington, DC: AAAI Press/MIT Press.

Christiansen, M. H., Hockema, S. A., & Onnis, L. (2006). Using phoneme distributions to discover words and lexical categories in unsegmented speech. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 172–177). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Christiansen, M. H., & Monaghan, P. (2006). Discovering verbs through multiple-cue integration. In R. M. Golinkoff & K. Hirsh-Pasek (Eds.), *Action meets word: How children learn verbs* (pp. 88–107). New York: Oxford University Press.

Christiansen, M. H., Onnis, L., & Hockema, S. A. (2008) *The secret is in the sound: From unsegmented speech to lexical categories.* Manuscript submitted for publication.

Clark, E. V. (1995). Later lexical development and word formation. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 393–412). Oxford, England: Basil Blackwell.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research, 22,* 109–131.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25,* 385–400.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71–99.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103,* 12203–12208.

Finch, S. P., Chater, N., & Redington, M. (1995). Acquiring syntactic information from distributional statistics. In J. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds.), *Connectionist models of memory and language* (pp. 229–242). London: UCL Press.

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development, 67,* 3192–3218.

French, R. M., Mermillod, M., Quinn, P., Chauvin, A., & Mareschal, D. (2002). The importance of starting blurry: Simulating improved basic-level category learning in infants due to weak visual acuity. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39,* 218–245.

Gallaway, C., & Richards, B. J. (1994). *Input and interaction in language acquisition*. Cambridge, England: Cambridge University Press.

Gauvain, J. L., Lamel, L. F., Adda, G., & Adda-Decker, M. (1993). Large vocabulary speech recognition in English and French. In *Proceedings of the IEEE Workshop on Automatic Speech Recognition*.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gerken, L. A. (1996). Prosody's role in language acquisition and adult parsing. *Journal of Psycholinguistic Research, 25,* 345–356.

Gerken, L. A., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological vs. syntactic phrases. *Cognition, 51,* 237–265.

Gerken, L. A., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32,* 249–268.

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

Golinkoff, R., Hirsh-Pasek, K., & Schweisguth, M. (2001). A reappraisal of young children's knowledge of grammatical morphemes. In J. Weissenborn & B. Hoele (Eds.), *Approaches to bootstrapping: Phonological, syntactic and neurological aspects of early language acquisition* (pp. 167–189). Amsterdam: Benjamins.

Green, T. R. G. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior, 18,* 481–496.

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13,* 318–324.

Hamasaki, N. (2002). The timing shift of two-year-olds' responses to caretakers' yes/no questions. In Y. Shirai et al. (Eds.), *Studies in language sciences (2)—Papers from the 2nd Annual Conference of the Japanese Society for Language Sciences* (pp. 193–206).

Harm, M., & Seidenberg, M. S. (1999). Reading acquisition, phonology, and dyslexia: Insights from a connectionist model. *Psychological Review, 106,* 491–528.

Harris, M. (1992). *Language experience and early language development. From input to uptake*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Harris, M., & Vincent, N. (1990). *The romance languages*. New York: Oxford University Press.

Harris, Z. (1946). From morpheme to utterance. *Language, 22*, 168–183.

Haskell, T. R., MacDonald, M. C., & Seidenberg, M. S. (2003). Language learning and innateness: Some implications of compounds research. *Cognitive Psychology, 47,* 119–163.

Hockema, S. A. (2006). Finding words in speech: An investigation of American English. *Language Learning and Development, 2,* 119–146.

Ishii, T. (2003). *The JUN corpus*, unpublished.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3,* 323–328.

Jusczyk, P. W., & Kemler-Nelson, D. G. (1996). Syntactic units, prosody, and psychological reality during infancy. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 389–408). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kelly, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language, 27,* 343–358.

Kemler-Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16,* 55–68.

Kempe, V., & Brooks, P. (2005). The role of diminutives in the acquisition of Russian gender: Can elements of child-directed speech aid in learning morphology? *Language Learning, 55,* 139–176.

Kuhl, P. K. (1999). Speech, language, and the brain: Innate preparation for learning. In M. D. Hauser & M. Konishi (Eds.), *The design of animal communication* (pp. 419–450). Cambridge, MA: MIT Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Maratsos, M., & Chalkley, M. (1980). The internal language of children's syntax. In K. E. Nelson (Ed.), *Children's language* (Vol. 2). New York: Gardner.

Mehler, J., Dommergues, J. Y., Frauenfelder, U. H., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20,* 298–305.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29,* 143–178.

Meier, R. P., & Bower, G. H. (1986). Semantic reference and phrasal grouping in the acquisition of a miniature phrase structure language. *Journal of Memory and Language, 25,* 492–505.

Mikheev, A. (1997). Automatic rule induction for unknown-word guessing. *Computational Linguistics, 23,* 405–423.

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

Mintz, T. H. (2004). Morphological segmentation in 15-month old infants. In A. Brugos, L. Micciulla, & C. E. Smith (Eds.), *Proceedings of the 28th Boston University Conference on Language Development Conference* (pp. 363–374). Somerville, MA: Cascadilla.

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., & Naka, N. (1998). *Wakachigaki Gaidorain WAKACHI98 v. 1.1* (Educational Psychology Forum Rep. No. FR–98–003). The Japanese Association of Educational Psychology.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential contribution of phonological and distributional cues in grammatical categorisation. *Cognition, 96,* 143–182.

Morais, J., Content, A., Cary, L., Mehler, J., & Segui, J. (1989). Syllabic segmentation and literacy. *Language and Cognitive Processes, 4,* 57–67.

Morgan. J. L. (1996). Prosody and the roots of parsing. *Language and Cognitive Processes, 11,* 69–106.

Morgan, J. L., & Demuth, K. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19,* 498–550.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of grammatical categories. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 263–283). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Murtagh, F. (1992). The multilayer perceptron for discriminant analysis: Two examples. In M. Schader (Ed.), *Analyzing and modeling data and knowledge* (pp. 305–314). New York: Springer-Verlag.

Nagata, M. (1999). A Part of Speech Estimation Method for Japanese unknown words using a statistical model of morphology and context. In *Proceedings of the 37th Annual Meeting of the Association of Computational Linguistics*, 277–284.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38,* 149.

Nelson, K. (1995). The dual category problem in the acquisition of action words. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 223–249). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE [An Internet derived lexical database of contemporary French]. *L'Année Psychologique, 101,* 447–462.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14,* 11–12.

Nielsen, H. F. (1989). *The Germanic languages*. Tuscaloosa: University of Alabama Press.

Oshima-Takane, Y., MacWhinney, B., Sirai, H., Miyata, S., & Naka, N. (Eds.). (1998). *CHILDES for Japanese* (2nd ed.). Japan: The JCHAT Project Nagoya, Chukyo University.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32,* 258–278.

Pallier, C., Christophe, A., & Mehler, J. (1997). Language-specific listening. *Trends in Cognitive Science, 1,* 129–132.

Peters, A. (1985). Language segmentation: Operating principles for the perception and analysis of language. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vols. 1–2, pp. 1029–1067). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: MIT Press.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Reali, F., Christiansen, M. H., & Monaghan, P. (2003). Phonological and distributional cues in syntax acquisition: Scaling up the connectionist approach to multiple-cue integration. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 970–975). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22,* 425–469.

Santelmann, L., & Jusczyk, P. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for processing limitations. *Cognition, 69,* 105–134.

Sapir, E. (1921). *Language*. New York: Harcourt Brace.

Sebastián-Gallés, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, 31,* 18–32.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Sereno, J. A., & Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech, 38,* 57–76.

Shady, M., & Gerken, L. A. (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language, 26,* 163–175.

Shi, R., Morgan, J., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language, 25,* 169–201.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 27,* B11–B21.

Shibatani, M. (Ed.). (1990). *The languages of Japan*. Cambridge, England: Cambridge University Press.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Reinhart & Winston.

Slobin, D. I. (Ed.). (1985). *The crosslinguistic study of language acquisition* (Vols. 1–2). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2,* 191–196.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4,* 156–163.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language, 27,* 71–86.

Valian, V., & Levitt, A. (1996). Prosody and adults' learning of syntactic structure. *Journal of Memory and Language, 35,* 497–516.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition, 24,* 744–755.

Wells, R. (1947). Immediate constituents. *Language, 23,* 81–117.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology, 50,* 509–535.