Using Prosodic Information to Discriminate between Function and Content Words

Jean-Marc Blanc & Peter F. Dominey

Institut des Sciences Cognitives
UMR 5015 CNRS - Université Claude Bernard Lyon 1
{blanc; dominey}@isc.cnrs.fr

Abstract

Early perceptual processing capabilities are likely to contribute to the categorization of lexical vs. grammatical words by newborns. This lexical categorization could be performed by detecting differences in the prosodic structure of these word categories. We demonstrated that this lexical categorization could be performed using many prosodic cues (duration, F0, energy and formants) automatically extracted for 10 different speakers.

1. Introduction

One of the most challenging questions in child language acquisition is how children learn syntax. A child must understand the relation between words in a sentence. A necessary prerequisite is that the language learner derives a knowledge of the different syntactic categories. But acquiring this knowledge presupposes the grammatical categories in terms of which they are defined; and the validity of grammatical categories depends on how far they support syntactic constraints.

Four main sources of information in linguistic input have been proposed as potentially useful in classifying lexical items into syntactic categories: Distributional Information [1], Semantic Bootstrapping [2], Phonological Constraints [3] and Prosodic Information [4].

The current research tests the hypothesis that prosody can be involved in the lexical categorization and thus in language acquisition. Our objective is to distinguish two lexical categories: **Content words** that have a meaning-related component such as nouns, verbs, adjectives, and adverbs, and **Function words** that are primarily structural, such as articles, prepositions, and auxiliaries.

2. Prosodic Foundations for Syntax

Jusczyk [5] proposed that infants are able to make immediate use of their sensitivity to prosodic markers as a means for organizing the input. We will now examine several investigations that have been realized to confirm this hypothesis.

2.1. Sensitivy to prosodic structure

2.1.1. Adults

Bagou et al. [6] evaluated the relative contribution of two prosodic cues, lengthening and F0 contour, in the processes of speech segmentation and storage of new words. Their results showed that prosodic information facilitates the acquisition of a new mini-language. Kelly [3] demonstrated that American subjects were able to exploit phonological cues to identify unknown words as verbs or nouns.

2.1.2. Children

Though children frequently fail to produce function morphemes in their earliest utterances, Gerken and McIntosh [7] have suggested that children by the age of 2 years have a representation of some specific function morphemes, and the context in which they appear. We will now consider the ability of newborns to exploit prosodic cues for the early basis of syntax.

2.1.3. Newborns

Since their first days of life, new-borns are also sensitive to intonational patterns. Two-month-olds discriminate multisyllabic stimuli that only differ on pitch and stress position [8]. Attention to prosodic characteristics at this age suggests that they may play an important role in how infants initially organise their knowledge about the sound structure of the native language

Especially, newborn infants are able to perceptually separate English word tokens into function/content categories [9]. At a later stage, such an ability could potentially help bootstrap infants into acquisition of grammar by allowing them to detect and represent classes of words on the basis of perceptible surface cues. However, the specific cues used by newborns to make this distinction have not been precisely determined.

2.2. Predictive prosodic cues for lexical discrimination

Shi et al. [4] have shown that multiple partly predictive cues to syntax categorization are available in speech. In English, functional items tend to have short syllable duration, low relative amplitude, whereas content words present the opposite patterns. Indeed, functional items universally tend to be productively and perceptually minimal.

Function words are often unstressed [10]. Mertens [11] have noted that in French, “if a grammatical word follows a content word then the content word is the end of an intonation group and receives a final accent”. In English, the content words are marked with a primary stress [12], which is indicated by increases in pitch.

Though stress is signalled in a complex, language dependant fashion, it is characterised by reinforcement in the articulatory energy that makes it salient at the auditory level. The physical variables, which carry that prosodic information, are F0, rhythm, and amplitude.

3. Simulation of syntax categorization

Shi et al. [4] investigated if various “presyntactic cues” (such as number of syllables, presence of a complex syllable

nucleus, presence of syllable coda, and syllable duration, to name only a few phonologically relevant cues) are sufficient to guide the assignment of words to rudimentary grammatical categories. Their investigation of English, Mandarin Chinese and Turkish shows that "sets of distributional, phonological, and acoustic cues distinguishing lexical and functional items are available in infant-directed speech across such typologically distinct languages as Mandarin and Turkish" [4]. Thus grammatical words tended to be acoustically and/or phonologically minimized in comparison to lexical words.

Durieux and Gillis [13] proposed an artificial learning system for lexical categorization with English (66.62% on CELEX lexical database) and Dutch (71.02% on INL lexical database) based on phonological and prosodic information.

The distinction between function and content words was also performed with a discriminant analysis on the 1000 most frequent words in the CHILDES corpus. The combination of 16 phonological cues allows 84.2% correct classification [14]. Simple Recurrent Networks were trained to predict the lexical category of the next input word from a corpus of child directed speech. These networks succeed to integrate 16 phonological cues with distributional information. The analysis of the hidden units allows 75.83% correct separation of nouns and verbs [15].

However in these simulation studies, a number of specific cues were extracted from the speech by a human expert. Here, we investigate whether cues automatically extracted from the contour of the fundamental frequency itself can be used to resolve the problem of lexical identification.

4. Material & Methods

4.1. Corpora

4.1.1. LSCP

This corpus contains 54 French sentences read by a single native speaker. The segmentation provided groups of words corresponding to adjacent words belonging to same category (content or function words; ~ 200 for each category [16]). Consonants, vowels and words were segmented by hand.

4.1.2. MULTEXT

Experiment 2 used French and English speech from the MULTEXT multilingual corpus developed for the study of prosody [17]. Stories were read by 20 different speakers (5 males and 5 females per language) which lead to a total of 8236 words for English, and 6945 words for French. Words were segmented by hand. For this corpus, correct classification results from the mean of the ten speakers of each language.

4.2. Prosodic information

We proposed to examine the contribution of prosody, with respect to four components*: intonation, loudness, voice quality,* and *rhythm*. Each of these components would be expressed in term of their physical parameters: *fundamental frequency*, *intensity*, *formants*, and the *duration* of vowels. We will now consider briefly this set of information and the way to extract it to model words or groups of words.

4.2.1. Duration

Each word or group of words was solely described by the mean duration of their vowels. In the LSCP corpus, consonants and vowels were segmented by hand, whereas for the MULTEXT corpus, they were automatically extracted with the algorithm developed by Pellegrinno. [18] We do not employ any normalization, as each speaker was studied individually. Furthermore the duration of words was also retained in a separate analysis.

4.2.2. Fundamental Frequency Contour

Fundamental frequency was obtained from the speech signal autocorrelation. A value of F0 was computed each 10ms. Each lexical group were represented by a combination of statistical parameters computed on the F0 values of a group of words:

- First value
- Last value
- Variation between first and last value
- Value of maximum
- Temporal position of maximum
- Duration of the group

4.2.3. Extended Prosodic Prototypes

Prosodic constituents are described as a function of time with a step of 10ms. Their values are obtained via the software PRAAT available online (http://www.praat.org).

Each word or group of words was described by eight dimensions: fundamental frequency (F0), intensity, the first three formants (F1, F2 and F3) and their bandwidth (ÄF1, ÄF2 and ÄF3). Each of these dimensions is encoded in vector of 15 statistic parameters:

- First value
- Last value
- Highest value
- Temporal position of maximum
- Lowest value (superior to 0)
- Temporal position of minimum
- Mean
- Standard deviation
- Ratio voiced versus unvoiced part. Unvoiced parts have zeros value in all dimensions.
- Number of rise and fall. (It corresponds to the number of times that the sign of the difference of two adjacent values change.)
- Number of rise and fall divided by duration
- Central moments of orders 2 to 5 (For order 2, central moment is proportional to variance, for order 3 to skewness and for order 4 to the kurtosis.)

Certain of these parameters are probably useless or redundant for our investigation. However, we were searching for the most complete description of the evolution of each dimension.

4.3. Learning algorithm

Three learning algorithms were retained to realize the lexical distinction using the different constituents of prosody. We employ a supervised statistical learning (discriminant analysis) and an unsupervised learning (self-organized maps with 5x5 units) to form categories without any correct exemplars. We also wanted to test whether these results could be obtained with a simple strategy, based on the distance

between two mean vectors, which were generated as prototypes of function and content categories. Each algorithm was trained on half of the data for each speaker, and then tested on the remaining half data.

5. Results

5.1. Duration

5.1.1. Vowel duration

The relevancy of vowel duration was assessed both by the method based on mean prototypes (71.5%) and by self-organised map (74.9%) on the LSCP corpus. This proves that French vowels are also influenced by the minimal character of function words.

We apply the algorithm of automatic segmentation on the 20 speakers of French and English. The duration of segments automatically classified as vowel serve as entry of a discriminant analysis. Lexical identification provides the same performance for French (73.3%) and English (73.3%).

5.1.2. Word duration

Shi et al. 1998 have proposed that both syllable duration and number of syllable contribute to the discrimination between function and content words. It implies that the duration of word is a relevant cue for this discrimination. In fact most of the groups of words were correctly identify (83.1% with mean prototype, and 84.5% using self-organized map). These results were replicated on the MULTEXT corpus (>80% for words of both languages, see first row of table 1).

We have shown that duration could be employed in lexical categorization, as it was pointed out by prosodic bootstrapping [4]. Is it possible to employ also fundamental frequency for lexical categorization?

5.2. Fundamental Frequency Contour

The set of parameters that led to the best performance (88.3% for the Kohonen map on the LSCP corpus) were variation between first and last value of the group of words, duration, final value and temporal position of maximum.

This experiment confirmed that the combination of different dimensions (duration and F0) contributes to lexical categorization, as it was already pointed out [4; 13; 15]. Could we extend these results to the MULTEXT corpus, and to new prosodic dimensions?

5.3. Extended Prosodic Prototypes

Duration and fundamental frequency could both be used for lexical categorization. Our objective is now the integration of every constituent of prosody (F0, duration, intensity and formants).

The following table shows percentage of correct identification of function and content items for French and English. One column indicates results for groups of words, as the corpus LSCP, whereas the other one show results for a single word. The first two rows are dedicated to a specific dimension: duration and then F0. However, some temporal information is still present in the representation of F0. The row '120 cues' represents the complete prosodic prototypes, with eight dimensions. The least row shows the results for the combination of duration and the 120 cues.

Table 1: *Correct classification of lexical item for the MULTEXT corpus.(G= groups of words; W= words)*

	English		French	
	G	W	G	W
Duration	77.3%	85.3%	72.3%	83.7%
F0	75.3%	80.9%	74.2%	82.3%
120 cues	79.1%	83.4%	79.5%	85.5%
Duration and all cues	79.1%	84%	79.9%	86.2%

First prosodic cues automatically extracted from the speech signal can be used to perform a lexical identification superior to 80% for 10 different speakers. Second, this result is available for French and English.

Duration is very predictive cue to lexical categorization, in particular when words are considered in isolation. F0 has a greater impact on lexical identification in the case of French. The results are equivalent between Duration and F0 for French, where as F0 gives inferior performance for English, in both Groups and Words condition. In every case, the adjunction of other prosodic cues to F0 increases performance. Meanwhile, performance was the highest for duration and English words. The combination of duration and the 120 cues reveals that duration brings supplementary information to the prosodic description.

Finally we concluded that regards to duration, prosodic cues allow greater performance, especially for French groups of same lexical type (+7%), less for French words and English groups (around 2%).

6. Discussion

6.1. Duration

Shi et al. [4] have demonstrated that syllable duration allow a classification of 64% of lexical item. We found superior results, even if vowels are automatically detected.

6.2. Fundamental Frequency Contour

We demonstrated that F0 contour could perform even better that duration, in the case of French groups of words. However, Shi et al. [4] shows that syllable duration was the best predictive cues to lexical categorization of function and content words, but they fail to obtain a significant difference with fundamental frequency for three different languages (English, Mandarin and Turkish). It could be due to the normalization for duration, realized by the mean of each syllable. Then if F0 have just a greater variation for one syllable of a content word, the variation will disappear during the mean on the entire word. Indeed, we already confirmed that F0 peaks, which last around a syllable, are valid cues for lexical categorization [19]. The results obtained for the MULTEXT corpus demonstrated that F0 contours provide same amount of information for lexical identification than duration.

6.3. Extended Prosodic Prototype

Our last experiment describes the combination of all prosodic dimensions (duration, F0, intensity and formants) to discriminate function from content words. We found that duration is a critical factor to this categorization, but that prosodic cues are useful for French groups of words of the

same lexical type, and in a certain extent for French words, and English groups of words. Linguistic literature have already emphasized that end of content group are marked by a final accent [11]. This suggests that units superior to words might be considered for syntax categorization.

6.4. Implication for the Acquisition of Language

In this context it would be of interest to study Child Directed Speech (CDS) that is known to include salient prosodic information. Indeed CDS displays an exaggeration of all prosodic parameters, in particular for F0 contour but also for duration and even in the location of formants. Thus prosodic cues to lexical discrimination would be more salient in CDS than in the speech corpus used in this study.

Furthermore, newborns are particularly sensitive to F0 contour and accents. Using the headturn preference procedure, Jusczyk, et al. [20] found that infants as young as 7.5 months are sensitive to a strong–weak stress pattern in bisyllabic words. In consequence, infants possess a mechanism that is sensitive to F0 contour. We already presented a neuro-inspired system (TRN) that can process F0 to identify prosodic attitudes [19], and also to discriminate function from content words with only F0 contour [21]. This study reveals that all prosodic components such as duration, amplitude and formants bring information for lexical categorization, could these dimensions be processed with the TRN?

7. Conclusions

These experiments demonstrated that duration, intonation contour, intensity and formants could contribute as a basis for an identification of Function vs. Content words that could bootstrap the acquisition of syntax. Furthermore we demonstrated that reliable information for lexical identification could be extracted from the prosody, whereas previous research [4, 13-15] hand-picked the variables to include in their simulations, based on adult knowledge and intuitions concerning relevant properties of lexical categorization. This observation has been demonstrated on two languages, English and French.

8. Acknowledgments

PFD is supported by the ACI Integrative and Computational Neuroscience, the OHLL and the Eurocores OMLL Projects. We thank F. Ramus, A. Christophe and E. Dupoux for insightful discussion and access to the LSCP corpus, and F. Pellegrinno for the segmentation in consonants and vowels of the MULTEXT corpus.

9. References

[1] Redington, M., Chater, N. & Finch, S., 1998. Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425-469.

[2] Pinker, S., 1984. Language learnability and language development. Cambridge, MA: Havard University Press.

[3] Kelly, M.H., 1995. The role of phonology in grammatical category assignments. In J.L. Morgan and K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (249-262). Mahwah, NJ: Lawrence Erlbaum Associates.

[4] Shi R., Morgan J.L., & Allopenna P., 1998. Phonological and acoustic bases for earliest grammatical category asssignment: a cross linguistic perspective. *Journal of child language,* 25, 169-201.

[5] Jusczyk, P.W., 1997. The discovery of spoken language. MIT Press.

[6] Bagou, O., Fougeron, C., & Frauenfelder, U.H., 2002. Contribution of Prosody to the Segmentation and Storage of "Words" in the Acquisition of a New Mini-Language, *Speech Prosody 2002*, 159-162.

[7] Gerken, L. A., & McIntosh, B. J., 1993. The interplay of function morphemes and prosody in early language. *Developmental Psychology*, 29, 448-457

[8] Spring, D. R. and Dale, P. S., 1977. Discrimination of linguistic stress in early infancy. In Journal of Speech and Hearing Research, 20, 224-232.

[9] Shi R., Werker J.F., & Morgan J.L., 1999. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words, *Cognition*, Volume 72, Issue 2, B11-B21.

[10] Gleitman, L.R. & Wanner, E., 1982. Language acquisition: The state of the state of the art. In E. Wanner & L.R. Gleitman (Eds.) *Language Acquisition: The State of the Art* (3-48). Cambridge University Press.

[11] Mertens, P., 1987. L'intonation du français. De la description linguistique à la reconnaissance automatique. Thèse de doctorat,Katholieke Universiteit Leuven.

[12] Hirst D., & Di Cristo A., 1998. A survey of intonation systems. in Hirst & Di Cristo (eds). *Intonation Systems : A Survey of Twenty Languages*.,1-44.

[13] Durieux, G., & Gillis, S., 2000. Predicting grammatical classes from phonological cues: An empirical test. In *Approaches to bootstrapping: phonological, syntactic and neurophysiological aspects of early language acquisition,* ed. B. Höhle, J. Weissenborn, 189-232. Amsterdam: Benjamins

[14] Monaghan, P., Chater, N. & Christiansen, M.C. (in preparation). Differential Contributions of Phonological and Distributional Cues in Language Acquisition.

[15] Reali, F., Christiansen, M.H., Monaghan, P., 2003. Phonological and Distributional Cues in Syntax Acquisition: Scaling-Up the Connectionist Approach to Multiple-Cue Integration. Proceedings of the 25th Annual Conference of the Cognitive Science Society.

[16] Ramus, F., Nespor, M., & Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.

[17] Campione, E., & Véronis, J., 1998. A multilingual prosodic database, *Proc. of ICSLP'98, Sidney*.

[18] Pellegrino, F.; Chauchat, J.-H.; Rakotomalala, R.; Farinas, J. Can Automatically Extracted Rhythmic Units Discriminate Among Languages? *Proc. of Speech Prosody 2002*.

[19] Blanc, J.M., Dominey, P.F., 2003. Identification of prosodic attitudes by a temporal recurrent network, *Cognitive Brain Research*, 17, 3, 693-699.

[20] Jusczyk, P. W., Houston, D., Newsome, M., 1999. The beginnings of word segmentation in English learning infants. *Cognitive Psychology*, 39, 159-207.

[21] Blanc, J.M., Dodane, C., Dominey, P.F., 2003. Temporal Processing for Syntax Acquisition: A simulation study. *Proc. of the 25th Annual Conference of the Cognitive Science Society*.