



Vocabulary structure and spoken-word recognition: Evidence from French reveals the source of embedding asymmetry

Anne Cutler^{1,2}, Laurence Bruggeman¹

¹ MARCS Institute, University of Western Sydney, Australia

² Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

a.cutler@uws.edu.au, l.bruggeman@uws.edu.au

Abstract

Vocabularies contain hundreds of thousands of words built from only a handful of phonemes, so that inevitably longer words tend to contain shorter ones. In many languages (but not all) such embedded words occur more often word-initially than word-finally, and this asymmetry, if present, has far-reaching consequences for spoken-word recognition. Prior research had ascribed the asymmetry to suffixing or to effects of stress (in particular, final syllables containing the vowel schwa). Analyses of the standard French vocabulary here reveal an effect of suffixing, as predicted by this account, and further analyses of an artificial variety of French reveal that extensive final schwa has an independent and additive effect in promoting the embedding asymmetry.

Index Terms: spoken-word recognition, vocabulary, lexical statistics, French, varieties

1. Introduction

The task of a spoken-word recogniser, be that a human listener or a machine, is to identify the word or words comprising an input from among all the possibilities afforded by the relevant vocabulary. Parts of this task are equivalent across languages. For instance, the vocabulary in all languages is huge and the phoneme repertoire is tiny. The average phoneme repertoire size across languages is 31, with the most common number of phonemes, as for instance in Spanish and Mandarin Chinese, being 25 [1]. Languages with more than 40 phonemes, such as British English, are by comparison profligate in populating the phoneme space; but compared with the hundreds of thousands of lexically represented units in a language's vocabulary, the variation in repertoire size is trivially inconsequential.

The need to make so many words out of so few building blocks means that inevitably words resemble one another (form minimal pairs such as *pat* versus *cat*, *put*, *pack*), overlap one another (as *pat* is found in *patch*, *pattern*, *patricide*, *spatter*, *expat*) and occur accidentally in sequences of words (*pat* in *keep at it*). So any speech input, in any language, will offer the human or machine recogniser multiple options for potential words or beginning paths through the incoming signal. This situation is unaffected by language phonology, word structure or other language-specific properties. All listening involves dealing with the simultaneous availability of multiple word candidates, and selecting between alternative interpretations, so that any model of spoken-word recognition by humans or of speech recognition by machine must confront this ambiguity resolution and candidate selection task.

Nevertheless, the task may differ subtly across languages in ways that affect both the reality of listening, and the factors that have to be taken into account in modelling it. Obviously, languages vary in the phonemes that make up their phonetic inventory, and in the words that comprise their vocabulary.

Such differences can have surprising effects. For instance, many phonetic inventories contrast /f/ and /s/, as in *leaf* vs. *lease*, but the way listeners distinguish these sounds differs as a function of whether their language has more other fricatives like /f/ or like /s/ [2]. So the very same speech information cueing the same contrast is processed differently depending on the rest of the inventory. Likewise whether or not a language has function words such as articles radically affects the frequency distribution of words that listeners encounter [3].

More serious for the listener are vocabulary properties that control the amount of competition involved in spoken-word recognition; that is, not only the population of alternative word candidates considered during speech processing, but the speed at which unwanted candidates can be rejected and the correct parse of the input determined. Phoneme repertoire size is a factor; the fewer the phonemes in a language, the longer words will tend to be, and the longer words are, the more embedding of shorter words they allow within them, and also the more embedding there will be just because it is more likely that any short sequence will indeed be an existing word. Comparisons between Spanish (with its 25-phoneme inventory) and British English (with its atypically large total of 44 phonemes) by [4] showed this indeed to be the case; the average length of Spanish words was roughly 20% above that for English words. When token counts were estimated by adjusting for frequency of occurrence of each word, the asymmetry became more pronounced: Spanish words contained on average 30% more phonemes than English words (and 40% more syllables).

Stress is another factor; both Spanish and English have stress, but they differ in whether listeners use suprasegmental stress cues (such as pitch, amplitude and timing of syllables) to distinguish between words. English listeners hardly use such cues (e.g., to distinguish the primary-stressed first syllable of *MUSIC* from the secondary stressed first syllable of *muSEUM* [5]). Spanish listeners, in contrast, make instant use of stress cues (e.g., *PRINcipe* and *prinCIpio* are distinguished in their first syllables [6]). Analyses in [4] revealed why; the amount of competition, i.e., embedded words, was hugely reduced by taking stress as well as segments into account in the Spanish computations, but adding stress insignificantly altered the English computations. English and Spanish listeners behave rationally, given the structure of their respective vocabularies.

Similar analyses of Dutch and German showed that, there too, attending to suprasegmental stress cues would reduce competition more rapidly than attending to segments only [7]. Thus despite their family relationship to English, both Dutch and German pattern more like Spanish than like English; and at least for Dutch, laboratory studies have shown word recognition by Dutch listeners to be more like that of Spanish listeners in this respect than like that of English listeners [8]. Indeed, not only do Dutch listeners use stress cues effectively in their own language, they also succeed in using stress cues in English that are ignored by English native listeners [5, 9].

Patterns of competition asymmetry are also dependent on the structure of vocabularies. In English, more embedding of shorter words in longer words occurs in initial than in final position. This is important for listeners, given that the number of competitors affects recognition [10] and competition effects differ at different word positions: Words with common onsets are easier to extract from preceding speech than words with rare onsets [11], but at the same time are recognized less rapidly than otherwise similar rare-onset words [12]; also, early- versus late-embedded words differ in their recognition patterns [13]. Listeners' response to competition at different points is, moreover, asymmetric. Competitors for word onsets play a stronger role than competitors for offsets – e.g., in eye-tracking studies, hearing *candle* induces early looks to a candy and to a handle, but the former tendency is stronger [14, 15]. All this suggests that initial competition has greater effects in listening than final competition, so that the asymmetric pattern of embedding in the vocabulary directly impacts the amount of competition that listeners are exposed to in real life.

This asymmetric pattern is not found in all languages, however; hardly any such asymmetry is observed in Japanese, for example [16]. On the other hand, Dutch and German show an asymmetry that is even greater than that in English. Comparison in [16] of the extent of the asymmetry in the three Germanic languages, combined with the Japanese statistics, suggested that the initial-final embedding asymmetry arises due to suffixing morphology and due to vowel reduction in final syllables, or to a combination of these factors.

Given that all of English, Dutch and German are stress languages, all have vowel reduction to schwa, and all have extensive suffixing, while Japanese has none of these features, the analyses of [16] could not arrive at a definitive conclusion as to the source of the embedding asymmetry. One of the above factors may have been crucial in supporting the asymmetry, with the rest being ancillary, or any one factor alone may be insufficient to bring the asymmetry about, and instead, the combination found in the three stress languages might be a necessary condition for its presence.

In the present study we extend this line of research to French, which crucially has some but not all of the features in question. Like Japanese, French is not a stress language; thus any pattern it shares with English, Dutch and German cannot depend on the presence of stress. Despite not having stress, French does have vowel reduction, so that any pattern it shares with Japanese cannot depend on reduced vowels. French also has inflectional morphology with many suffixes, so that any effect that arises from such morphology should be expected for French as for English, Dutch and German, with all together then differing from Japanese. As well, French has a phoneme inventory that is large (39 phonemes), so that effects resulting from phoneme inventory size should group French closer to English than to Spanish. Thus analyses of embedding patterns in French will furnish us with a more precise understanding of the role of morphology, the presence of stress, and phoneme inventory size in the provision of competition in listening.

2. Lexical analyses of French

For statistical analysis of the French vocabulary we used the Lexique lexical database [17], specifically Lexique 3.72 (down-loaded June 8, 2012). Lexique 3.72 contains 142694 lexical entries, with associated frequency counts based on both written and spoken sources (in the case of the written frequencies a corpus of 15 million words).

For comparability with the analyses performed for English, Spanish, Dutch, German and Japanese [4, 7, 16], we pruned the database by removing phrasal entries (any item containing an internal space in its name), all proper nouns (including the names of letters of the alphabet), and any entry with a pronunciation identical to some earlier entry (i.e., homophones were accorded one entry only). This produced a core lexicon of 70327 words, highly comparable in size to the lexical databases used in the analyses of English and Spanish (approximately 70000 and 60000 words respectively [4]).

Also to test for cross-analysis comparability, we compute for French the word length and embedding computations done for English and Spanish in [4]. Length computations have so far only been done for English and Spanish (not for Japanese, Dutch or German). We predict that, solely on the grounds of phoneme inventory size, French word length results will resemble results from English (a Germanic language) more than results from Spanish (like French, a Romance language).

2.1. Word length

We first computed word length in phonemes. The analyses in [4] had compared a vocabulary based on 44 phonemes, with an average word length of 6.94 phonemes/2.72 syllables, to one based on 25 phonemes which had a mean word length of 8.3 phonemes/3.48 syllables. French, with 22 consonants and 17 vowels [18], is predicted to resemble the former case more than the latter, and indeed it does. The mean word length in phonemes is 6.83 and the mean length in syllables is 2.9.

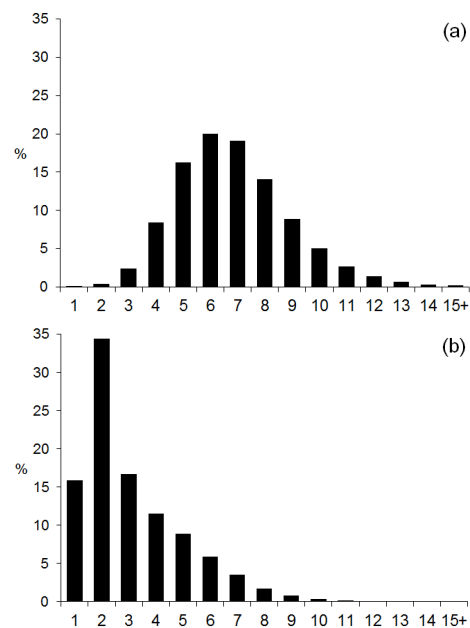


Figure 1: *Distribution of word types in the French vocabulary (a) and likely proportion of word tokens in speech samples (b), by word length in phonemes (1 to 15 or more).*

Adjusting the values for word frequency gives an estimate of token occurrence in the corpora providing the frequency counts; given that shorter words occur more frequently, the token-based mean length is of course lower, at 3.13 phonemes and 1.44 syllables. Again these values are very close to the reported values for English in [4] (3.5/1.43). Figures 1a and 1b display the French type and token length distributions, again strongly resembling the distribution figures for English in [4].

2.2. Embedded words in the vocabulary

Table 1 shows the proportion of words in the vocabulary containing some other existing word embedded within them, sorted by length of the carrier, length of the embedded word, and syllable position at which the embedded word begins. As in all other languages, embedding in the French vocabulary is rife. The overall count of embedded words is 207995, or 2.96 embedded words per carrier word, comparable with every other language for which such counts have been made.

Table 1. *Proportion of polysyllabic carrier words (CW) with embedded words (EW) of different lengths*

# syll CW	# syll EW	Location EW Onset in CW				
		1	2	3	4	5
2	1	.904	.695			
3	1	.887	.839	.741		
	2	.438	.313			
4	1	.905	.865	.803	.748	
	2	.403	.233	.313		
	3	.254	.175			
5	1	.913	.853	.897	.784	.718
	2	.415	.214	.233	.305	
	3	.136	.074	.217		
	4	.189	.129			

Having established that the word length and embedding statistics in the present analyses pattern as predicted, we can use the French statistics to shed new light on questions that the earlier analyses of other languages could not fully answer.

2.3. Location of embedded words

The cross-language comparison by [16] described in the introduction concerned the asymmetry of embedding in word-initial versus word-final position. This asymmetry was first established for English [4, 19], and proved in many laboratory studies of spoken-word recognition to underlie variation in competition patterns and hence variation in speed and accuracy of word recognition [13-15]. In Japanese, however, the asymmetry was hardly to be seen, as Figure 2 shows.

As described, this could be for one of several reasons. Japanese does not have the suffixal morphology typical of English, whereby the ends of words are likely to be suffixes and thus less likely to be, or to be part of, embedded words. Also, English is a stress language with a marked preference for word-initial stress [20], so that final syllables are more likely to be unstressed and thereby less likely to be viable stand-alone words. Either stress or morphology could account for such a difference between English and Japanese.

The analyses of [16] could not rule out either of these candidate explanations, and in fact they suggested that both factors may play a role, and contribute separately. Dutch showed a skew that was even more asymmetric than English, and German a yet greater one. Both Dutch and German are suffixing stress languages, like their relative English, but they both have more extensive suffixing than English (e.g., a wider range of verbal inflections and of plural forms). They also both have a lesser tendency to vowel reduction in all unstressed syllables than English does [7], with more reduced vowels in final syllables than in earlier-occurring positions.

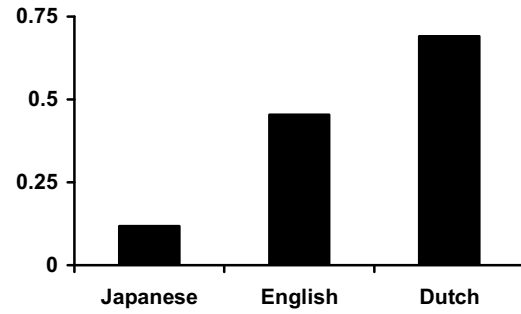


Figure 2: *Asymmetry of initial versus final embedding in the vocabulary of Japanese, English and Dutch (added ratio of initially embedded over finally embedded words), from [16].*

In short, either morphology or stress alone could thus still account for the cross-linguistic difference of Japanese versus the stress languages, or it could be the case that both factors together are required. An analysis of French, however, offers us the opportunity to seek a more complete answer. French, like the Germanic languages, has extensive suffix morphology, but is not a stress language. If suffixing alone is responsible for the initial-final embedding asymmetry, then French will show an asymmetry at least equal to that of English, perhaps greater, given that French (like Dutch and German) has more different inflections than English. However, if stress (solely or in combination) lies at the root of the asymmetry, then French should show no significant asymmetry at all, like Japanese. Finally, if there is substance to the suggestion in [16] that morphology and stress both contribute, independently, to asymmetric embedding in a vocabulary, then we should find French (with only one of these features) patterning somewhere in between Japanese (with neither) and English (with both).

We first calculated the number of embedded words aligned with the onset of their carrier word (initial embeddings) versus the number aligned with the offset (final embeddings), and computed the resulting initial/final ratio in the same manner as for the languages analysed in [16]. The left column of Figure 3 shows the result: the ratio for French is 1.265:1, an asymmetry of .265, falling almost exactly midway between the values for Japanese (1.118:1) and English (1.454:1) shown in Figure 1.

This suggests therefore that suffixing morphology indeed is able to make a separate contribution to the likelihood of a language having more embedded words word-initially than word-finally (with all the consequences of this for the listener's task in spoken-word recognition). The fact that French fell between Japanese and the stress languages indicates that both stress and suffixing contribute substantially to the asymmetry.

The larger asymmetry in Dutch than in English, and the even larger asymmetry in German, were explained in [16] as a result of those two stress languages having more final syllables containing schwa. Many Dutch plurals and verbal inflections end in *-en*, and German has many words ending with a syllable containing schwa (the difference between German and either English or Dutch can be readily seen in cognate sets; *cat*, *lamp*, *guitar* in English are *kat*, *lamp*, *gitaar* in Dutch, but *Katze*, *Lampe*, *Gitarre* in German, with in each case the final German syllable consisting of the vowel schwa). Our next analysis explores effects of final schwa, and is made possible by schwa being indeed a vowel in the French repertoire.

3. Discussion

When speech recognisers, human or otherwise, process natural speech, their principal task is to detect the words that are indeed present in the signal and reject all words that may be partially or fully supported, but only by accident. Decades of research on human spoken-word recognition have shown that native languages lead their users to develop listening strategies that deal efficiently with the speech input they are likely to receive [24], and these include expectation-based ways of dealing with embedded word competitors [11-15].

Phonology and word structure differ across languages, and have considerable implications for the amount and position of embedding [4, 16] and hence for the competition that arises in listening and for the type of strategies that listeners are led to develop. In particular the greater likelihood of initial than of final embedding has implications for listeners' behaviour [14, 15]. Prior comparisons of embedding frequency and position across languages [4, 7] had shown such initial/final asymmetry to be present in English and also in Spanish, Dutch and German. However, Japanese shows no such asymmetry. More extensive analyses of Japanese versus the Germanic languages indicated that the asymmetry appeared when languages had suffixes and/or words with schwa in the final syllable. Though all of English, Dutch and German have predominantly initial stress and thus have many unstressed final syllables, the fact that German has very many more schwa-final words (such as *Katze* etc.) was proposed to underlie the very much greater size of the asymmetry in German.

Our present work with French has enabled us to confirm that both suffixing (which French has) and extensive final-syllable schwa (which, in the standard variety, it does not have) play a role in the initial/final embedding asymmetry. Moreover, we could confirm that these two factors are indeed independent and additive, by adding the final-schwa effect in a created version of French. Our analyses lead us to conclude that the initial/final embedding asymmetry is jointly caused by suffixes and syllables with final schwa. Japanese, which has none of these, has no significant asymmetry; either or both effects induce the asymmetry, and the more pronounced they are, the more pronounced the asymmetry too.

The present analyses would not have been possible in any of the other languages examined. It is not possible to compare the effect of suffixing and of schwa-final syllables separately in the stress languages, because they already have both features to varying degrees. It is also not possible to compare them in Japanese or in Spanish, since neither language allows final-schwa syllables as a phonological possibility or indeed has schwa in the phoneme inventory. Creation of a schwa-final version of either language would have simply created illegal sequences. French, however, both has schwa in its phoneme repertoire, and provides a phonologically legal reason for it to occur in the required position. Hence French has brought us significantly closer to definitively answering the question of what lies at the root of the initial/final embedding asymmetry.

4. Acknowledgements

This research was funded by the Max Planck Institute for Psycholinguistics, NWO-SPINOZA ("Native and Non-native Listening", AC), and by a MARCS Institute doctoral scholarship (LB). Preparation of the paper at the University of Canterbury, NZ, was supported by an Erskine Fellowship to AC. We thank Romain Fiasson for helpful discussions.

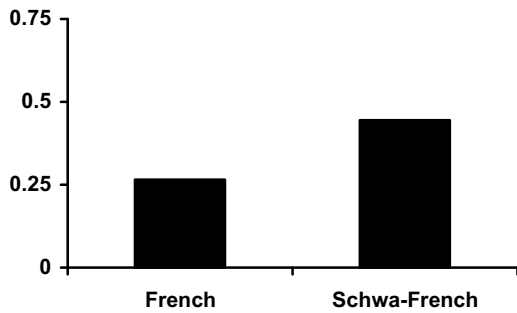


Figure 3: *Asymmetry of initial versus final embedding in the vocabulary of French (left column: added ratio of initially embedded over finally embedded words), and in an artificially constructed version of French in which final syllables in words such as *bonne* are pronounced with schwa (right column).*

2.4. Exploring the effect of final schwa

What is the effect of final schwa in a non-stress language vocabulary? French, while it does not have stress, does have the reduced vowel schwa. In standard French many potential opportunities for pronunciation of schwa, e.g., in affixes, are in fact silent, but as these opportunities are realised in some varieties of French [21-23], we may regard the potential for schwa pronunciation to be underlying in French in general.

To further test the role of final schwa in the ratio of initial to final embedding, we created an artificial version of French in which schwa was added to the lexical pronunciation of all potential syllables in final position that are silent in standard French but realised in some varieties; thus *bonne* would have two syllables and *petite* three, in each case with the last vowel being schwa. This entirely imaginary but phonologically possible variant of French we termed "schwa-French".

We then calculated the resulting embedding patterns in the vocabulary. Overall, the pattern was just as in the standard vocabulary, at fractionally under three embedded words per word of the vocabulary. But as Figure 3 shows, the initial/final asymmetry changed markedly, to an English-like 1.445:1. So if a language readily allows final syllables with schwa (that are unlikely to be stand-alone words), it will have relatively fewer final embeddings and hence a greater initial/final asymmetry.

2.5. Vocabulary mismatch and embedding

As an addendum, since we now had two versions of the French lexicon, we also explored the outcome of effectively presenting tokens from one lexicon to the other. Speakers of one variety of a language converse all the time with speakers of other varieties. Does vocabulary mismatch affect amount and position of available competition in listening?

For a listener with standard French hearing schwa-French input, the overall number of competitors drops slightly to 2.54 per word, but the initial/final embedding asymmetry rises to 1.43:1, similar to that of schwa-French. For a listener with a schwa-French lexicon hearing standard input, the overall number of embeddings rises a little (to 3.17 per word), but the initial/final embedding asymmetry rises steeply, to 1.85:1. Listening strategies formed from embedding expectations in the native variety would thus be misleading for such input.

5. References

- [1] I. Maddieson, *Patterns of Sounds*. Cambridge University Press, Cambridge, 1984.
- [2] A. Wagner, M. Ernestus, and A. Cutler, "Formant transitions in fricative identification: The role of native fricative inventory," *Journal of the Acoustical Society of America*, vol. 120, pp. 2267-2277, 2006.
- [3] R. Shi, J. L. Morgan, and P. Allopenna, "Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective," *Journal of Child Language*, vol. 25, pp. 169-201, 1998.
- [4] A. Cutler, D. Norris, and N. Sebastián-Gallés, "Phonemic repertoire and similarity within the vocabulary," in *Proceedings of the 8th International Conference on Spoken Language Processing*, 2004, pp. 65-68.
- [5] N. Cooper, A. Cutler, and R. Wales, "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," *Language and Speech*, vol. 45, pp. 207-228, 2002.
- [6] S. Soto-Faraco, N. Sebastián-Gallés, and A. Cutler, "Segmental and suprasegmental mismatch in lexical access," *Journal of Memory and Language*, vol. 45, pp. 412-432, 2001.
- [7] A. Cutler and D. Pasveer, "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition," in *Proceedings of Speech Prosody*, 2006, pp. 237-400.
- [8] W. v. Donselaar, M. Koster, and A. Cutler, "Exploring the role of lexical stress in lexical recognition," *Quarterly Journal of Experimental Psychology Section A*, vol. 58, pp. 251-273, 2005.
- [9] A. Cutler, "Greater sensitivity to prosodic goodness in non-native than in native listeners," *Journal of the Acoustical Society of America*, vol. 125, pp. 3522-3525, 2009.
- [10] D. Norris, J. M. McQueen, and A. Cutler, "Competition and segmentation in spoken-word recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 21, pp. 1209-1228, 1995.
- [11] A. H. v. d. Lugt, "The use of sequential probabilities in the segmentation of speech," *Perception & Psychophysics*, vol. 63, pp. 811-823, 2001.
- [12] M. S. Vitevitch, "Influence of onset density on spoken-word recognition," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 28, pp. 270-278, 2002.
- [13] M. S. Cluff and P. A. Luce, "Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 16, pp. 551-563, 1990.
- [14] P. D. Allopenna, J. S. Magnuson, and M. K. Tanenhaus, "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models," *Journal of Memory and Language*, vol. 38, pp. 419-439, 1998.
- [15] J. M. McQueen and M. C. Viebahn, "Tracking recognition of spoken words by tracking looks to printed words," *Quarterly Journal of Experimental Psychology*, vol. 60, pp. 661-671, 2007.
- [16] A. Cutler, T. Otake, and L. Bruggeman, "Phonologically determined asymmetries in vocabulary structure across languages," *Journal of the Acoustical Society of America*, vol. 132, pp. EL155-EL160, 2012.
- [17] B. New, C. Pallier, M. Brysbaert, and L. Ferrand, "Lexique 2: A new French lexical database," *Behavior Research Methods*, vol. 36, pp. 516-524, 2004.
- [18] International Phonetic Association, *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*: Cambridge University Press, Cambridge, 1999.
- [19] J. M. McQueen and A. Cutler, "Words within words: Lexical statistics and lexical access," in *Proceedings of the Second International Conference on Spoken Language Processing*, 1992, pp. 221-224.
- [20] A. Cutler and D. M. Carter, "The predominance of strong initial syllables in the English vocabulary," *Computer Speech & Language*, vol. 2, pp. 133-142, 1987.
- [21] F. Dell, *Les règles et les sons: introduction à la phonologie générative* vol. 25: Hermann, 1973.
- [22] J. Durand, "Alternances vocaliques en français du midi et phonologie du gouvernement," *Lingua*, vol. 95, pp. 27-50, 1995.
- [23] J. Durand, C. Slater, and H. Wise, "Observations on schwa in southern French," *Linguistics*, vol. 25, pp. 983-1004, 1987.
- [24] A. Cutler, *Native listening: Language experience and the recognition of spoken words*: MIT Press, Cambridge MA, 2012.