

FINDING SCHWA: COMPARING THE RESULTS OF AN AUTOMATIC ALIGNER WITH HUMAN JUDGMENTS WHEN IDENTIFYING SCHWA IN A CORPUS OF SPOKEN FRENCH

Peter M Milne

Dept. of Linguistics, University of Ottawa, 70 Laurier Ave East., ON, Canada, K1N 6N5, pmiln099@uottawa.ca

1 INTRODUCTION

The value in working with natural language corpora is the ability to collect large volumes of empirical data with which to test research hypotheses. The challenge is to generate these data quickly and accurately. Accumulating the linguistic data required to test and evaluate hypotheses can be a time consuming and labour intensive job. The availability of an accurate automatic aligner, which can aid in the analysis of larger volumes of natural language data than is possible when working manually, would greatly assist in this task. This paper addresses the question of how the results of an automatic aligner compare with human judgments when identifying the presence or absence of schwa in a word-final, post-consonantal context.

All word-final, postconsonantal possible schwa insertion sites were identified in the standardized read text portion of investigations conducted in both Québec and France. All data was systematically coded for presence or absence of schwa by trained researchers. The data was also time aligned at both the word and phone level by a forced aligner. The results of the two methods of coding were statistically compared to determine their level of agreement. Results show a significant correlation between the two methods and a high likelihood of overall agreement. Possible effects of dialect or phonetic context were investigated using a two-way, between subjects analysis of variance. Initial results have found no significant effect of dialect or context. An examination of the confusion matrix indicated that the majority of errors were the result of the aligner finding schwa where none existed. This suggests that the results of automatic alignment are comparable with human judgments in both dialects of French, regardless of phonetic context, but that improvements to the acoustic model for French schwa are needed.

1.1 The forced aligner

The forced aligner used is the Penn Phonetics Lab Forced Aligner (Yuan and Liberman, 2008), modified for use with French (described in Milne, 2010). The Penn Phonetics Lab Forced Aligner is an automatic phonetic alignment toolkit based on HTK, the Hidden Markov Model toolkit maintained by the Cambridge University Engineering department. Both are distributed for use free of charge. It takes as input an audio .wav file along with a corresponding .txt orthographic text file and, using the HTK toolkit, produces a Praat TextGrid with interval boundaries for segments and words on two tiers. The pronunciation dictionary used in conjunction with the modified aligner is an expanded version of Lexique, version 3 (New et al.,

2001, 2004). Lexique is a database that provides > 135,000 words of French including orthographic and phonemic representations, syllabification, parts of speech, gender and number, frequency, and associated lemmas, etc. This information is stored in tables that can be downloaded or searched online. Lexique is an open database to which everyone is encouraged to participate. Using a word list generated from the transcription files, the pronunciation dictionary was expanded to include alternate pronunciations for all words containing possible schwa insertion sites, as well as liaison consonants and consonant cluster reduction possibilities. The current expanded dictionary used for this project contains entries for each word in the transcription containing a possible schwa insertion site having pronunciation candidates both with and without schwa.

In addition to expanding the pronunciation dictionary, the aligner was modified by mapping all French phones that do not occur in English onto the acoustic models that best matched their phonetic qualities. These mappings included the nasalized vowels [œ̃, ɛ̃, ɑ̃, ɔ̃], the rounded vowels [ø, œ, y], and the consonant [ɥ].

1.2 The corpus

The corpus used was obtained through the *Projet Français Contemporain* (Durand et al., 2002, 2005). This is an ongoing research project aimed at providing researchers interested in the French language with a database of oral data. One of the main goals is to gather data from as many varieties of French as possible, in all parts of the world, in order to investigate dialectal variation. It is the largest collection of spoken French, and one of the largest collections of spoken language data for any language.

The data obtained was from the standardized read text portion of 9 investigations conducted in both France and Québec. This data set is appropriate for several reasons. First, the text was systematically constructed to maximize both the frequency of possible schwa sites as well as the variety of contexts under which schwa patterns can be tested. Second, there exists a comparable volume of data for both dialects under consideration.

2 METHODS

The corpus contains data from 38 speakers (19 each from Québec and France) and includes 6458 possible schwa insertion sites. Each site was individually coded (manually by researchers trained in the PFC protocol and automatically via the forced aligner) for presence or absence of schwa, position in the word, and both left- and right-hand contexts.

From this, 3160 word-final, post-consonantal possible schwa insertion sites were extracted. The results of automatic alignment were compared with human judgment with respect to the presence or absence of schwa. An agreement value was calculated with values of 1, indicating agreement between the two methods of coding on the presence or absence of schwa, or 2, indicating disagreement.

876 sites were randomly sampled from the dataset. These were evenly balanced for Dialect ($n=438$) and RHContext ($n_{\text{vowel}}=149$, $n_{\text{consonant}}=145$, $n_{\text{pause}}=144$). An overall accuracy rate was calculated by dividing the number of items that agreed on the presence or absence of schwa by the total number of items. This accuracy rate was compared with both chance and a baseline scenario in which schwa was never inserted. In order to determine whether the values for agreement differed according to either dialect or right-hand context, this sample was subjected to a two-way, between subjects analysis of variance with the dependent variable as Agree, and two independent variables of Dialect (2 levels, “QF”, “FR”) and RHContext (3 levels, “V”, “C”, “pause”). Finally, the conditions under which the two methods of coding disagreed were examined.

3 RESULTS

In our sample, 651 of 876 items agreed between the two methods of coding for an overall accuracy rate of 74.3%. This is significantly better than chance ($\chi^2(1)=110$, $p<.05$), but not better than the baseline scenario in which no schwa was ever inserted. In fact, the results indicate that the baseline scenario would agree with the PFC coding significantly more often than the forced aligner ($\chi^2(1)=119$, $p<.05$).

The results of the ANOVA indicate no significant differences in agreement either between dialects ($F(1,872)=0.15$, $p=0.67$), or among contexts ($F(2,872)=0.41$, $p=0.66$). This suggests that the forced aligner produces judgments similar to the manual results of the PFC on the presence or absence of schwa in both dialects and in all three of the right-hand contexts.

In examining the conditions under which the two methods of coding disagree, there were 225 instances of disagreement between the two methods of coding. 218 of those instances, or 96.9%, were of the kind where the forced aligner judged a schwa segment to be present, while the human PFC coder perceived no schwa segment. This is shown below in Table 1. In Table 1, **FA** indicates the values obtained for items by the forced aligner and **PFC** indicates the values obtained for items by manual coding.

Table 1: Confusion Matrix

	PFC schwa	PFC no schwa
FA schwa	50	218
FA no schwa	7	601

4 DISCUSSION AND CONCLUSIONS

This study was conducted to determine whether the Penn Phonetics Lab Forced Aligner, modified for use with French, could aid in automating the task of determining the presence or absence of schwa in a large corpus of spoken French. The results suggest that the modified aligner works equally well in both dialects, and is unaffected by right-hand context. However, the disagreements between the two methods of coding indicate that the modified aligner frequently finds schwa where no schwa exists. It is suspected that this confusion may be due to the differences in acoustic cues between the French and English schwa. The Penn Phonetics Lab Forced Aligner is built on English monophone acoustic models and the French schwa is more front and rounded than the English schwa (closer to [ø] or [œ]). Perhaps in this context (word-final, post-consonantal), the aligner is interpreting some elements of consonantal release as an inserted schwa. A possible way to improve this aligner is to update the acoustic model of schwa to better approximate the different phonetic qualities of French schwa as compared with English schwa.

REFERENCES

- Durand, J., Laks, B., and Lyche, C. (2002). La phonologie du français contemporain: usages, variétés et structure. In Pusch, C. and Raible, W., editors, *Romanistische Korpuslinguistik - Korpora und gesprochene Sprache/Romance Corpus Linguistics Corpora and Spoken Language*, pages 93–106. Gunter Narr, Tübingen.
- Durand, J., Laks, B., and Lyche, C. (2005). Un corpus numérisé pour la phonologie du français. In Williams, G., editor, *La linguistique de corpus*, pages 205–217. Presses Universitaires de Rennes, Rennes.
- Milne, P. (2010). The effects of syllable position on allophonic variation in québec french /ʁ/: A corpus analysis using a modified version of the penn phonetics lab forced aligner. <http://aix2.uottawa.ca/~pmiln099>.
- New, B., Pallier, C., Brysbaert, M., and Ferrand, L. (2004). Lexique 2 : A new french lexical database. *Behavior Research Methods, Instruments, and Computers*, 36(3):515–524.
- New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: Lexique. *L'Année Psychologique*, 101:447–462. <http://www.lexique.org>.
- Yuan, J. and Liberman, M. (2008). Speaker identification on the scotus corpus. In *Proceedings of Acoustics '08*.