Speech Prosody of French Regional Varieties

Mathieu Avanzi [1], *Nicolas Obin* [2], *Alice Bardiaux* [3], *Guri Bordal* [4, 5]

[1]Chaire de linguistique française, Université de Neuchâtel, Switzerland;
[2]IRCAM-CNRS UMR 9912-STMS, Paris, France; [3]FNRS, Université catholique de Louvain, Belgium; [4]University of Oslo; [5]MoDyCo, UMR 7114, Université Paris Ouest Nanterre, France
mathieu.avanzi@unine.ch, nobin@ircam.fr, alice.bardiaux@uclouvain.be, guri.bordal@ilos.uio.no

Abstract

This paper compares the prosody of 6 varieties of French spoken in three different areas: France (Paris and Lyon), Belgium (Tournai and Liège), and Switzerland (Geneva and Neuchâtel). The objective is to adress whether some regional varieties, namely those of Geneva and Tournai, are closer to standard French (i.e. the varieties spoken in France, represented here by Paris and Lyon) than others (Neuchâtel and Liège). The recordings of the same text read by 4 speakers representing each variety were semi-automatically processed in order to study accentuation, speech rate and rhythm, and 8 prosodic measures that can possibly discriminate the 6 varieties were compared. A top-down clustering supported evidence for the expected classification with regard to the "standard" varieties, while a bottom-up clustering pointed out a more contrasted configuration.

Index Terms: Prosody, regional French, standard French, Accentual Phrase, articulation rate, speech rate.

1. Introduction

In this paper, we assess to which extent "regional" varieties of French differ from "standard" varieties with respect to their speech prosody. We compare 6 varieties of European French spoken in three countries (France, Switzerland and Belgium), chosen because they are thought to represent cardinal and median varieties of a "regionality" scale (see Figure 1):

- [FR-ST] includes to the varieties of French spoken in Paris (FR-75) and Lyon (FR-69), considered as "standard" varieties;
- [FR+] includes French spoken in Geneva (SW-GE) and Tournai (BE-TO), considered *weakly* regionally marked varieties of Switzerland and Belgium since they are very close to the French border;
- [FR-] includes French spoken in Neuchâtel (SW-NE) and Liège (BE-LI), considered as *strongly* regionally marked varieties since they are pretty far from the French border;

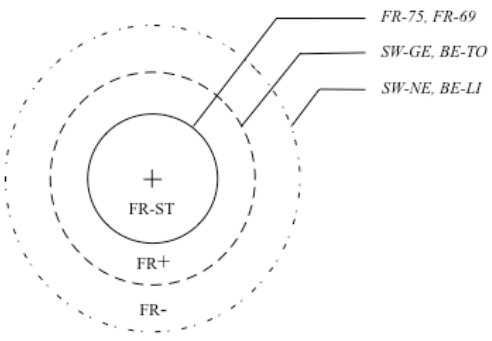

Figure 1: *Regionality scale of French varieties*

A large number of prosodic features can be used to address the prosodic similarity of French regional varieties with regard to their proximity to the standard varieties. Regarding rhythm and accentuation, the observations reported in existing studies of variation in European French prosody (see [1]-[10]) led us to formulate the following two hypotheses:

(H1) Speakers of European "regional" varieties speak more slowly than speakers of "standard" varieties;

(H2) Speakers of European "regional" varieties tend to mark the penultimate syllable of accentual phrases with a prominence, while speakers of standard varieties mark the last syllable of accentual phrases with a prominent syllable.

Very little work has been done to address the empirical grounding of these two hypotheses. For (H1), the only existing studies focusing on articulation and speech rate are concerned with Paris ([7] and [9]) and different oïl dialectal varieties compared with a regional variety (Neuchâtel) and a standard variety (Nyon) [10]. These studies resulted into conflicting conclusions. [10] pointed out significant differences between the "standard" and the regional varieties, while [7] and [9] did not. Concerning (H2), [10] compared Swiss speakers (speakers from Nyon) and Belgian speakers (speakers from Tournai, Liège and Gembloux, the former being a [FR-] variety on our scale) with speakers from France, and pointed out that the penultimate and final syllables of inter-pause groups tended to be longer in the Belgian and Swiss varieties than in the French varieties. This paper presents a corpus-based study of French prosody regional variations. The mentioned hypotheses will be specifically addressed, and the productions of speakers in six regional varieties will be studied and compared. While the study of regional variations primarily aims to improve the theoretical modeling of the French prosodic system, this may also be used to improve automatic speech recognition and synthesis systems [11].

2. Speech Material

The speech material was collected from the "Phonologie du Français Contemporain" (PFC) speech database, which contains speech productions of thousands of speakers from French-speaking areas all around the world [4]. For each of the 6 varieties studied in this paper (see Figure 1), we selected the recordings of the same text read by 4 speakers (two male and two female, two young speakers, 20-30 and two older speakers, 40-50). The text contains 398 words which are phrased in 22 sentences, and is 130 seconds long on average. In all, the entire corpus is 52 minutes long.

2.1. Linguistic transcription

Speech samples were transcribed orthographically in Praat [12], and automatically aligned with the script Easyalign [13],

which provides a 3-layer segmentation in phones (transcribed with SAMPA), syllables and words. The automatic alignments were then manually corrected. Prominent syllables and disfluencies (for instance, hesitation or syntactic interruptions) were independently identified on perceptual bases by two of the authors, following a procedure initiated by [14]. Kappa statistics indicated substantial inter-annotator agreement (κ = 0.65). Finally, the reference tier that will be further used was obtained as follows: syllables which present agreement were defined as the reference. In cases of disagreement, a third expert (one of the authors) determined the prominent status (+/- prominent) of the syllable. The text was also parsed in Accentual Phrase (henceforth AP): a clitic group (one content word and its dependent functional words, see [16]) right bounded by a prominence syllable in the reference tier was considered as the head of an AP.

2.2. Acoustic Measures

To address (H1) (rhythm hypothesis), five measures were compared:

articulation rate: syllable rate of each AP included in the sentence, excluding silent pauses;

speech rate: syllable rate of all the sentences, including silent pauses;

accentuation rate: proportion of prominent syllables of the APs included in the sentence;

AP weight: number of syllables in each AP;

(%V, ΔC): proportion of vocalic segments and standard deviation of inter-vocalic segment duration over the sentence [17].

To address (H2) (accentuation hypothesis), two measures were compared:

lengthening: duration of the final syllable of an AP compared to the preceding syllable;

F0 rise: difference in semitones of the F0 peak of the voiced region of the final syllable of the AP compared to the semitones of the F0 of the preceding syllable. The F0 peak was determined by the F0 value that was maximally distant from the average F0 of the voiced region of the syllable – either positive (F0 rise) or negative (F0 fall).

Finally, each measure was determined and locally averaged – when necessary – over the sentence. Hence, each speaker is represented by a distribution of characteristics over the 22 sentences of the speech database.

2.3. Robust statistics

In order to provide the characteristics of each variety, taking into consideration possible outliers or speakers with markedly different characteristics from the other speakers of their variety, the conventional average characteristics (mean μ and standard deviation σ) were determined using a robust estimation assuming a normal distribution of the characteristics considered.

$$\overline{\mu}_X = \text{median}(\mathbf{x})$$
$$\overline{\sigma}_X = 0.7413 \times \text{iqr}(\mathbf{x})$$

where: median(.) and iqr(.) denote the median and interquartile range, and $\mathbf{x}$ the vector of the observed characteristics.

Additionally, the average characteristics of a variety were determined by the pooled characteristics of all the speakers of the variety, and not by the speaker averages. This strategy was adopted in order to ensure robust post-hoc analysis (one-way ANOVA) that could then be used to assess significant differences between the varieties. Indeed, the number of observations per speaker is generally significantly greater than the number of speakers of a variety. Thus, statistics of observations are more robust than those of speakers – the number of observations for each speaker being roughly equal.

3. Results

Two strategies were adopted to adress our hypotheses: a top-down clustering, in which the 6 varieties were grouped *a priori* in 3 classes according to the expected classification (§3.1), and a bottom-up clustering, in which no *a priori* classification was made to determine the similarities of the varieties (§3.2). The top-down clustering was used to assess whether the expected classification is consistent with acoustic measures (inter-group variations), whereas the bottom-up clustering was used to assess whether the expected classification could be automatically retrieved from acoustic measures (intra-group variations). For both strategies, the clustering of each variety/group was determined using an agglomerative hierarchical clustering method [18], in which the varieties/groups are iteratively clustered by pairs according to the distance of their mean characteristics. Additionally, post-hoc analysis (one-way ANOVA) was used to assess significant differences within and between the clusters obtained.

In the following, the figures present clustering obtained for some study-case characteristics that will be used for discussion. Significant differences are shown by means of a color representation. The groups with a uniform color indicate that there are no significant differences within the group, while a change of color indicates significant differences for each pair of varieties in the group. The significance threshold was set at a 99% significance confidence level (*p-value*<0.01).

3.1. Top-down clustering

In the top-down clustering, the 6 varieties were grouped in three classes according to their position in the regionality scale (see above, Figure 1). Table 1 gives the mean values and the standard deviation of the 8 prosodic features calculated for each of the three classes. According to H1 and H2, [FR+] and [FR-] varieties are expected to be closer to each other than to the standard varieties [FR-ST], while [FR+] varieties are expected to be closer to the [FR-ST] varieties than the [FR-] varieties. The evidence from top-down clustering supports the expected classification in the majority of the cases – with the exception of the %V characteristic. Two configurations were observed:

1. *Expected classification*: [FR-ST] vs. [FR+ and FR-], with [FR+] closer to [FR-ST] than [FR-]

This configuration was observed for all of the characteristics with the exception of the %V. Significant differences were observed between [FR-ST] and the other groups. Additionally, a significant difference was observed between [FR+] and [FR-] when it comes to articulation rate, speech rate, accentuation rate, ΔC and F0 rise, while no significant differences were observed for the other measures.

2. *Consistent classification*: [FR-ST and FR+] vs. [FR-], with [FR+] closer to [FR-] than [FR-ST]

This configuration was observed for the %V measure, with a significant difference between [FR-ST and FR+] and [FR-]. Further, no significant difference was observed between [FR+] and [FR-] (with a significance threshold of 99%), while a significant difference was observed with the 95% significance threshold.

	FR-ST (FR-69, FR-75)		FR+ (SW-GE, BE-TO)		FR- (SW-NE,BE-LI)	
Hyp. I: rhythm features						
articulation rate	6.1	(0.5)	5.6	(0.6)	5.3	(0.5)
speech rate	5.3	(0.5)	4.7	(0.6)	4.4	(0.5)
accentuation rate	36.2	(4.9)	39.5	(5.7)	40.0	(5.4)
AP weight	3.4	(0.6)	3.4	(0.6)	3.2	(0.6)
ΔC (x100)	4.0	(0.7)	4.8	(0.8)	5.2	(0.9)
%V	46.1	(5.7)	45.3	(4.5)	48.1	(5.0)
Hyp. II: accentuation features						
lengthening	1.59	(0.24)	1.62	(0.26)	1.64	(0.21)
F0 rise	-1.0	(1.7)	-0.2	(1.2)	0.3	(1.0)

Table 1. *Means and standard deviations for the 3 expected groups.*

As an illustration, Figure 2 shows the clustering obtained for the articulation rate.

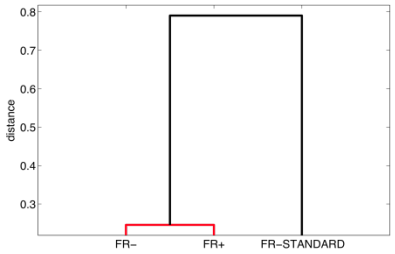

Figure 2: *Top-down clustering obtained for articulation rate*

3.2. Bottom-up clustering

In the bottom-up clustering, the varieties are not grouped according to their position on the regionality scale (see Figure 1 above). Table 2 below gives the mean values and the standard deviation calculated for each of the 8 prosodic features calculated for each variety. Compared with the top-down classification, bottom-up clustering reveals a more sharply contrasted situation; the differences between the 6 varieties are not as systematic as (H1) and (H2) would lead one to expect. Three tendencies were observed:

1. *Expected classification*: the obtained clustering matches the expected classification.

This configuration was observed for the articulation rate and the ΔC measures, with a significant difference between all groups (Figure 3).

2. *Consistent classification*: the obtained clustering is consistent but does not match the expected classification.

This configuration was observed for the speech rate, accentuation rate, AP weight, and F0 rise – with variable configurations and significant differences. For instance, Figure 4 presents the clustering obtained by the speech rate measure, in which a significant difference is observed between the [FR-ST] varieties and the other varieties, while the other varieties form a uniform group; there is no significant distinction between [FR+] and [FR-]. This suggests that the articulation rate is more accurate than speech rate for the description of regional variations. Interestingly, the F0 rise measure presents a configuration in which the clustering is clearly driven more by the geographic background than by distance from the standard variety (Figure 5).

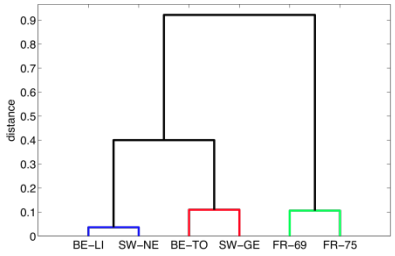

Figure 3: *Bottom-up clustering obtained for articulation rate*

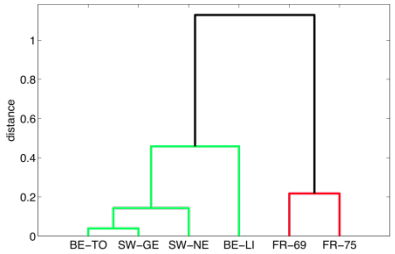

Figure 4: *Bottom-up clustering obtained for speech rate*

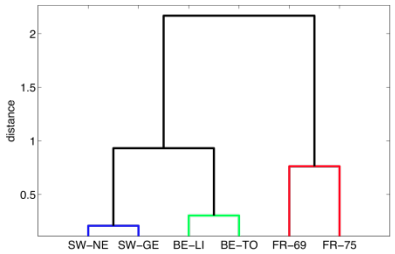

Figure 5: Bottom-up clustering obtained for F0 rise.

	FR-69		FR-75		BE-LI		BE-TO		SW-NE		SW-GE	
Hyp. I: rhythm features												
articulation rate	6.2	(0.4)	6.1	(0.5)	5.3	(0.5)	5.6	(0.6)	5.3	(0.5)	5.5	(0.5)
speech rate	5.4	(0.4)	5.2	(0.5)	4.2	(0.6)	4.7	(0.7)	4.5	(0.4)	4.8	(0.5)
accentuation rate	36.8	(4.8)	35.0	(5.1)	39.4	(5.7)	40.3	(6.0)	41.8	(5.7)	38.8	(5.0)
AP weight	3.4	(0.4)	3.5	(0.7)	3.2	(0.5)	3.4	(0.5)	3.1	(0.5)	3.2	(0.5)
ΔC (x100)	4.0	(0.6)	4.1	(0.7)	5.2	(0.8)	4.7	(0.7)	5.3	(0.9)	4.9	(0.8)
%V	44.7	(4.7)	48.4	(5.8)	46.9	(5.0)	45.2	(4.2)	48.8	(4.8)	46.2	(5.1)
Hyp. II: accentuation features												
lengthening	1.57	(0.21)	1.67	(0.25)	1.71	(0.24)	1.51	(0.22)	1.62	(0.20)	1.72	(0.28)
F0 rise	-0.5	(1.0)	-1.4	(2.7)	-0.9	(1.1)	-0.6	(1.2)	0.7	(1.2)	0.9	(1.0)

Table 2. *Mean and standard deviation for the 6 varieties.*

3. *Non-consistent clustering*

This configuration was observed for the %V measure solely – which means that %V is not a reliable measure to describe regional variations. Hence, variations in speech rhythm are related to consonantal variations but not to vocalic reduction. This indicates that standard varieties have a more regular syllable structure than other varieties.

3.3. Discussion

Depending on the method used for clustering, top-down (§3.1) or bottom-up (§3.2), the 8 prosodic measures vary in relevancy for the description of regional variations. Table 3 presents a summary of the prosodic measures that are significant in both conditions (++), the measures that are significant in all but one condition (+), and the measures that are not significant (-).

Hyp. I: rhythm features	
articulation rate	++
speech rate	+
accentuation rate	+
AP weight	-
ΔC (x100)	++
%V	-
Hyp. II: accentuation features	
lengthening	-
F0 rise	+

Table 3. *Measures reliability for the description of regional variations.*

Two features conduct to the expected classification of the 6 varieties according to (H1) in both conditions, namely articulation rate and ΔC; while speech rate and accentuation rate correspond with the expected classification in the top-down clustering solely; finally, lengthening, AP weight and %V are not good predictors to distinguish between the varieties studied here. If none of these features confirms the hypotheses (H2) in all conditions, this may be explained by the averaging of salient characteristics that are observed in specific conditions only. In particular, [2] indicated that regional accentuation features are observed only for some specific APs. Thus, the statistical averaging of a speaker's characteristics may mask some significant but not systematic differences. While, it is difficult to compare studies that have not been conducted with the same material and the same methodology, our results confirm that speakers referred as "regional varieties" tend to speak more slowly than speakers of "standard varieties" and this is in fact mainly due to two different cues: articulation rate and ΔC (the latter indicating that standard varieties have a more regular syllable structure than other varieties). On the contrary, our results do not support the claim of [10], who found significant differences between the lengthening of the penultimate syllables of the prosodic groups of speakers of a "regional variety" and speakers of a "standard variety".

4. Conclusion

This study presented a corpus-based study of French prosody regional variations and addressed the prosodic similarity of French regional varieties with regard to their proximity to the standard varieties. Robust stastistics were proposed to estimate characteristics of the considered varieties, and clustering methods were introduced to cluster the varieties with respect to their average characteristics. Top-down clustering supports evidence for the expected classification while bottom-up clustering pointed out a more contrasted configuration. Further studies will introduce fine grain prosodic measures considered as relevant for the description of French regional variations. For instance, [9] pointed out that AP final pitch rises in the Swiss productions were mostly anchored on the penultimate syllable, while in Parisian French, AP final pitch rises were generally anchored on the final syllable.

5. References

[1] Andreassen, H., Maître, R., Racine, I. "La Suisse", in Detey, S., J. Durand et al. [eds], Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement, Paris/Gap, Ophrys, 201-212, 2010.

[2] Carton, F. "A la recherche d'intonations régionales", Proc. CILPR, 249-257, 1986.

[3] Carton, F., Espesser, R., Vaissière, J. "Etude sur la perception de l'accent régional du Nord et de l'Est de la France", Proc. 12th ICPhS, 422-425, 1991.

[4] Durand, J., Laks, B., Lyche, C. (eds), Phonologie, variation et accents du français, Paris, Hermès, http://www.projet-pfc.net/, 2009.

[5] Hambye, Ph. & Simon, A.C. "The variation of pronunciation in Belgian French: from segmental phonology to prosody", in Guess, R., Lyche, C. & Meisenburg, T. (eds). Phonological Variation in French: Illustrations from Three Continents, Amsterdam, John Benjamins, to appear.

[6] Knecht, P., Rubattel, C. "A propos de la dimension sociolinguistique du français en Suisse romande", Le Français moderne, 52, 138-150, 1984.

[7] Schwab, S., Racine, I. "Le débit lent des Suisses romands : mythe ou réalité ? ", submitted.

[8] Simon, A. C. (éd.). La variation prosodique régionale en français, Paris, Hermès, to appear.

[9] Sterling Miller, J. Swiss French prosody. Intonation, rate, and speaking style in the Vaud Canton, PhD, University of Illinois, 2007.

[10] Woerhling, C. Accents régionaux en français. Perception, analyse et modélisation à partir de grands corpus, PhD, Université Paris Sud, 2008.

[11] Obin, N. MeLos: Analysis and Modelling of Speech Prosody and Speaking Style, PhD, Ircam-UPMC, Paris, 2011.

[12] Boersma, P. & Weenink, D. "Praat: doing phonetics by computer (Version 5.5)". www.praat.org, 2011.

[13] Goldman, J.-Ph. "EasyAlign: an automatic phonetic alignment tool under Praat", Proc. Interspeech, 3233-3236, http://latlcui.unige.ch/phonetique/, 2011.

[14] Avanzi, M., Simon, A. C., Goldman, J.-Ph., Auchlin, A. "C-PROM. An annotated corpus for French prominence studies". Proc. Prosodic Prominence: Speech Prosody Workshop, 2010.

[15] Cohen, J., "A coefficient of agreement for nominal scales", Educational and Psychological Measurement, 20, 37-46, 1960.

[1] Jun, S. A., Fougeron, C., Realizations of accentual phrase in French intonation, Probus, 14, pp. 147-172, 2002.

[16] Garde, P. L'accent, Paris, PUF, 1965.

[17] Ramus, F., Nespor, M. & Mehler, J. "Correlates of linguistic rhythm in the speech signal", Cognition, 73/3, 265-292, 1999.

[18] Trevor H., Tibshirani, R. Friedman, J. Hierarchical clustering. The Elements of Statistical Learning, New York, Springer, 520-528, 2009.