

La variation prosodique dialectale en français. Données et hypothèses

Mathieu Avanzi¹ Nicolas Obin² Guri Bordal^{3,4} Alice Bardiaux⁵

(1) Chaire de linguistique française, Université de Neuchâtel, Suisse

(2) IRCAM-CNRS UMR 9912-STMS, Paris, France

(3) Université d'Oslo, Norvège

(4) MoDyCo, UMR 7114, Université Paris Ouest Nanterre, France

(5) FNRS, Université catholique de Louvain, Belgique

mathieu.avanzi@unine.ch, nobin@ircam.fr, guri.bordal@ilos.uio.no,
alice.bardiaux@uclouvain.be

RÉSUMÉ

Dans cet article, nous comparons la prosodie de 6 variétés de français parlées en France (Paris et Lyon), en Belgique (Tournai et Liège) et en Suisse (Genève et Neuchâtel). L'objectif est de voir si les 6 variétés considérées peuvent être discriminées sur la base de critères exclusivement prosodiques. Les enregistrements du même texte lu par 4 locuteurs pour chacune des variétés sont transcrits, alignés et codés pour l'étude de l'accentuation, du phrasé et du rythme. Les résultats d'une méthode de classification non-supervisée guidée par les hypothèses (*top-down*) aboutissent à des résultats cohérents avec une classification *a priori* des variétés sur une échelle d'éloignement dialectal, alors qu'une méthode de classification non-supervisée émergente (*bottom-up*) donne lieu à des résultats plus contrastés.

ABSTRACT

Speech Prosody of Dialectal French: Data and Hypotheses

This paper contrasts 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 address whether some prosodic criteria can help to classify distinct dialectal varieties from each other. 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. 8 prosodic measures that can possibly discriminate the 6 varieties were compared. A top-down clustering supports evidence for the expected classification, while a bottom-up clustering points out a more contrasted configuration.

MOTS-CLÉS : Prosodie, français dialectal, accentuation, rythme, débit.

KEYWORDS : Prosody, dialectal French, accentuation, rhythm, speech rate.

1 Introduction

Dans cet article, nous proposons une méthodologie en vue d'évaluer la distance qui sépare plusieurs variétés dialectales de français au regard de leurs propriétés prosodiques uniquement. En pratique, nous comparons 6 variétés de français parlées en France, Belgique et Suisse, sélectionnées parce qu'elles représentent les points cardinaux d'une échelle d'éloignement dialectal (cf. figure 1) : le français parlé à Paris (FR-76) et le français parlé à Lyon (FR-69), qui constituent dans la littérature les variétés dites de

référence ou « standard » (désormais [FR-ST]) ; les variétés de français parlées à Genève (FR-GE) et Tournai (FR-TO), désormais [FR+], considérées comme des variétés peu marquées par rapport aux variétés françaises susmentionnées ; et les variétés de français parlées à Neuchâtel (FR-NE) et à Liège (FR-LI), désormais [FR-], qui peuvent être décrites comme des variétés suisses et belges fortement marquées par rapport à celles de Paris et de Lyon :

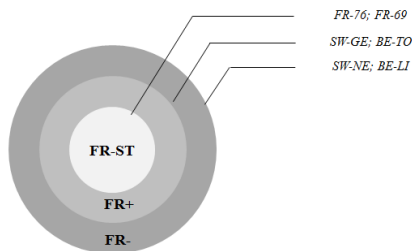


FIGURE 1 – Echelle d'éloignement dialectal de quelques variétés de français parlées en France, en Belgique et en Suisse.

Un grand nombre de mesures prosodiques pourraient être calculées en vue de rendre compte de la similarité ou de l'éloignement prosodique existant entre différentes variétés dialectales du français. En ce qui concerne l'accentuation, le rythme et le débit, un examen de la littérature portant sur la variation prosodique du français européen (*cf.* références bibliographiques *infra*) nous conduit à formuler les hypothèses suivantes :

- (H1) Les locuteurs francophones de Suisse et de Belgique parlent plus lentement que les sujets pratiquant une variété hexagonale de la zone d'oïl;
- (H2) Les locuteurs francophones de Suisse et de Belgique auraient tendance à marquer la syllabe pénultième de leurs groupes accentuels par une proéminence, alors que les sujets parlant une variété hexagonale de la zone d'oïl se contentent de marquer prosodiquement la dernière syllabe de leurs groupes accentuels uniquement.

À notre connaissance, peu d'auteurs ont cherché à tester empiriquement la validité scientifique de telles hypothèses. Concernant H1, les seules études existantes comparent le débit de locuteurs parisiens (Schwab & Racine, à par. et Sterling Miller, 2007) et de locuteurs originaires de divers sites de la zone d'oïl avec celui de locuteurs vivant à Neuchâtel ou dans la région de Nyon (Woerhling, 2008). Les conclusions de ces études livrent des résultats pour le moins contradictoires : alors que Woerhling (2008) souligne l'existence de différences significatives entre les variétés d'oïl et les variétés suisses, Schwab & Racine (à par.) et Sterling Miller (2007) concluent à l'absence de différence. Concernant H2, Woerhling (2008) compare des locuteurs français de la zone d'oïl avec des locuteurs belges (de Gembloux, de Tournai et de Liège) et suisses (de Nyon). En se concentrant sur les syllabes pénultièmes et finales des groupes inter-pausaux identifiés dans les enregistrements examinés, l'auteur montre que de telles syllabes ont tendance à être plus longues dans les variétés belges et suisses que dans les variétés françaises. Du travail reste donc à faire pour mieux comprendre ce qui distingue différentes variétés dialectales de français parlé en Europe. La recherche décrite dans cet article se propose d'enrichir notre connaissance de la prosodie des variétés dialectales de français. Les

hypothèses mentionnées ci-dessus seront spécifiquement abordées à travers l'étude comparée de productions de locuteurs francophones originaires de 6 villes françaises, belges et suisses.

2 Données

Les données sur lesquelles nous avons travaillé ont été collectées dans le cadre du projet Phonologie du Français Contemporain (PFC, cf. Durand, Laks & Lyche, 2009), qui constitue une base de données contenant des enregistrements de centaines de locuteurs originaires de toute la francophonie. Pour chacune des variétés considérées dans cet article, nous avons sélectionné le même texte lu par 4 locuteurs (deux hommes, deux femmes, deux locuteurs jeunes entre 20 et 30 ans, et deux locuteurs plus âgés, entre 40 et 50 ans). Le texte contient 398 mots regroupés en 22 phrases, et dure en moyenne 130 secondes. Au total, notre corpus d'étude est d'une durée de 52 minutes. Dans un premier temps, chacun des 24 enregistrements a été segmenté en phrases graphiques et transcrit en orthographe standard dans Praat (Boersma & Weenink, 2012), puis aligné en phonèmes, syllabes et mots graphiques à l'aide du script EasyAlign (Goldman, 2011). Les alignements ont été corrigés manuellement. Les proéminences accentuelles et les disfluences (segments associés à une hésitation ou un piétinement sur l'axe syntagmatique) ont été codées par deux experts (deux des auteurs) parallèlement, suivant pour cela une procédure proposée par Avanzi, Simon, Goldman & Auchlin (2010). Une tire de comparaison a ensuite été générée et l'accord mesuré. Cet accord ayant été jugé substantiel ($\kappa = 0,65$), un troisième expert (un des auteurs) a tranché dans les cas de discordance pour aboutir à un codage de référence. Enfin, un des auteurs a identifié dans une tire dédiée les groupes clitiques dont le bord droit étaient associé à une proéminence, segmentant ainsi le texte en syntagmes accentuels (désormais SA, Jun & Fougeron, 2002).

En vue de tester H1, hypothèse selon laquelle plus on s'éloigne du centre de l'échelle d'éloignement dialectal, plus on parle lentement, nous avons calculé et comparé les 5 mesures suivantes :

- Taux d'articulation : nombre de syllabes/seconde pour chacun des SA sans inclure les pauses ;
- Débit : nombre de syllabes/seconde pour chacun des énoncés, incluant les pauses ;
- Densité accentuelle : proportion des syllabes proéminentes pour chacun des SA contenu par énoncé ;
- Poids métrique du SA : nombre de syllabes par SA ;
- (%V ; ΔC) : proportion de segments vocaliques et écart-type de la durée des segments intervocaliques à l'intérieur d'un même énoncé.

En vue de tester H2, hypothèse selon laquelle plus on s'éloigne du centre de l'échelle d'éloignement dialectal, plus on a tendance à rendre saillante la syllabe pénultième des groupes prosodiques, deux mesures ont été prises et comparées :

- Allongement pénultième : différence de durée entre la syllabe finale et la syllabe pénultième du SA ;
- Pente de Fo : différence en demi-tons entre la valeur du pic de Fo de la partie

voisée de syllabe finale du SA et à la valeur moyenne de Fo en demi-tons de la syllabe précédente. La valeur du pic de Fo est la valeur la plus éloignée de la Fo moyenne de la partie voisée de la syllabe – elle est soit positive (montée) soit négative (descente).

Finalement, chaque mesure a été déterminée et relativisée, quand cela était nécessaire, par rapport à l’empan contextuel que constitue l’énoncé. Ainsi, chaque locuteur du corpus est représenté par la distribution de ces caractéristiques par rapport à l’ensemble des 22 énoncés de notre base de données. Afin de donner une description des caractéristiques prosodiques de chacune des variétés de notre corpus en fonction des mesures acoustiques décrites ci-dessus, tout en tenant compte des valeurs aberrantes ou singulières par rapport aux moyennes calculées pour chacune des variétés, les caractéristiques moyennes conventionnelles (moyenne μ et écart-type σ) ont été déterminées sur les bases d’une estimation statistique robuste supposant une distribution normale des caractéristiques considérées, selon la formule suivante :

$$\begin{aligned}\bar{\mu}_X &= \text{median}(\mathbf{x}) \\ \bar{\sigma}_X &= 0.7413 \times \text{iqr}(\mathbf{x})\end{aligned}$$

Où $\text{median}(\cdot)$ et $\text{iqr}(\cdot)$ désignent respectivement la médiane et l’écart interquartile ; \mathbf{x} étant le vecteur des caractéristiques observées. De plus, les caractéristiques moyennes d’une variété ont été déterminées par les caractéristiques mises en commun de tous les locuteurs de la variété, et non sur la base des moyennes des locuteurs. Cette stratégie a été adoptée afin d’assurer la robustesse des analyses post-hoc (ANOVAs à un facteur) utilisées par la suite en vue d’évaluer les différences significatives entre les variétés. En effet, le nombre d’observations par locuteur est généralement nettement supérieur au nombre de locuteurs d’une variété. En procédant de cette manière, les statistiques d’observations sont plus robustes que celles couramment utilisées dans la littérature – le nombre d’observations pour chaque locuteur étant à peu près égal.

3 Résultats

Afin de tester les deux hypothèses présentées ci-dessus, deux stratégies distinctes ont été adoptées. Dans un premier temps, nous avons opté pour une approche guidée par les hypothèses, ou *top-down* (§ 3.1.) : dans cette approche, les 6 variétés ont été regroupées a priori en trois groupes selon la classification prévue par l’échelle de d’éloignement dialectal (*cf.* figure 1 *supra*). Dans un second temps, nous avons opté pour une approche émergente, ou *bottom-up* (§ 3.2.), ne préjugant en rien de la nature des regroupements possibles entre variétés. Alors que la première approche a été choisie en vue d’évaluer si le classement attendu dépeint par la figure 1 est conforme aux mesures acoustiques calculées (variation inter-variétés), la méthode émergente a été utilisée pour évaluer si des mesures acoustiques *stricto sensu* permettaient de retomber sur la classification attendue (variation intra-variété). Pour les deux stratégies, le regroupement de chaque variété a été déterminé en utilisant une méthode classification hiérarchique agglomérative (Trévor *et al.*, 2009) : les variétés/groupes sont confrontés par paires itérativement en fonction de la distance de leurs caractéristiques moyennes. En outre, des analyses post-hoc (simples ANOVAs) ont été utilisées pour évaluer l’existence de différences significatives au sein et entre les groupes. Dans les pages qui suivent, les figures présentent les regroupements obtenus pour des exemples précis qui seront utilisés

pour la discussion. Les différences significatives sont indiquées grâce à un code de couleurs : si le regroupement est de couleur uniforme, cela veut dire qu'il n'y a pas de différence significative au sein du groupe, tandis qu'un changement de couleur indique des différences significatives pour chaque paire de variétés dans le groupe. Le seuil de significativité a été fixé à un niveau de confiance portée à 99% ($p < 0,01$).

3.1 Approche guidée par les hypothèses

Dans la configuration guidée par les hypothèses, les 6 variétés ont été regroupées en trois catégories *a priori*, selon leur position sur l'échelle d'éloignement dialectal (cf. figure 1 *supra*). Le tableau 1 ci-dessous donne les valeurs moyennes et l'écart-type des 8 caractéristiques prosodiques calculées pour chacune des trois classes :

	FR-ST (FR-69, FR-75)		FR+ (SW-GE, BE-TO)		FR- (SW-NE, BE-LI)	
Hyp. I: mesures de rythme et de débit						
Vitesse d'articulation	6.1	(0.5)	5.6	(0.6)	5.3	(0.5)
Débit	5.3	(0.5)	4.7	(0.6)	4.4	(0.5)
Densité accentuelle	36.2	(4.9)	39.5	(5.7)	40.0	(5.4)
Poids métrique du SA	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: mesures d'accentuation						
Allongement	1.59	(0.24)	1.62	(0.26)	1.64	(0.21)
Pente de Fo	-1.0	(1.7)	-0.2	(1.2)	0.3	(1.0)

TABLE 1 – Moyenne et variation standard pour les 3 classes de variétés.

Globalement, nos résultats révèlent des classifications conformes à nos hypothèses. Sur les 8 mesures calculées, 7 permettent de conclure que les variétés [FR+] sont plus proches de la variété [FR-ST] que de la variété [FR-]. En outre, une différence significative a été observée entre [FR+] et [FR-] en ce qui concerne la vitesse d'articulation, le débit de parole, le taux d'accentuation, le ΔC et la pente de Fo, alors qu'aucune différence significative n'a été observée pour les autres mesures (poids métrique du SA et allongement pénultième). La 8e mesure, la mesure de %V, aboutit à une classification pour laquelle les variétés [FR+] sont plus proches de la variété [FR-] que de la variété [FR-ST]. Aucune différence significative n'a été observée entre [FR+] et [FR-] avec un seuil de significativité de 99%.

3.2 Approche émergente

Dans la configuration émergente, les variétés ne sont pas regroupées en fonction de leur position sur l'échelle d'éloignement dialectal (voir figure 1 ci-dessus). Le tableau 2 ci-après donne les valeurs moyennes et l'écart-type des 8 caractéristiques prosodiques calculées pour chacune des 6 variétés étudiées dans cet article.

Les résultats obtenus avec l'approche émergente ne convergent avec le classement attendu qu'avec 2 des 8 paramètres envisagés : la vitesse d'articulation et la mesure de ΔC , avec des différences significatives entre tous les groupes. La figure 2 ci-après donne une illustration de cette situation. Des paramètres tels que le débit, le taux d'accentuation, le poids métrique du SA, l'allongement pénultième et la pente de Fo aboutissent à un *classement cohérent*, mais ne correspond pas à la classification prévue. Les configurations de significativité sont variables, allant de « aucune différence » à « des

différences plus ou moins prévisibles » (point non développé ici). Par exemple, la figure 3 (à gauche) présente le regroupement obtenu pour la mesure de débit, dans lequel une différence significative n'est observée qu'entre les variétés [FR-ST] et les autres variétés, tandis que les autres variétés forment un groupe homogène, à l'intérieur duquel il n'y a pas de distinctions significatives. Cela suggère que la vitesse d'articulation constitue un prédicteur plus précis que le débit pour la description des variations régionales. Fait intéressant, la mesure de la pente de Fo présente une configuration dans laquelle le regroupement est clairement plus motivé davantage par l'origine géographique que par la distance avec la variété standard (cf. figure 3, à droite).

	FR-69		FR-75		BE-LI		BE-TO		SW-NE		SW-GE	
Hyp. I: mesures de rythme et de débit												
Vitesse d'articulation	6.2	(0.4)	6.1	(0.5)	5.3	(0.5)	5.6	(0.6)	5.3	(0.5)	5.5	(0.5)
Débit	5.4	(0.4)	5.2	(0.5)	4.2	(0.6)	4.7	(0.7)	4.5	(0.4)	4.8	(0.5)
Densité accentuelle	36.8	(4.8)	35.0	(5.1)	39.4	(5.7)	40.3	(6.0)	41.8	(5.7)	38.8	(5.0)
Poids métrique du SA	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: mesures d'accentuation												
Allongement	1.57	(0.21)	1.67	(0.25)	1.71	(0.24)	1.51	(0.22)	1.62	(0.20)	1.72	(0.28)
Pente de Fo	-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 – Moyenne et écart-type pour les 6 variétés.

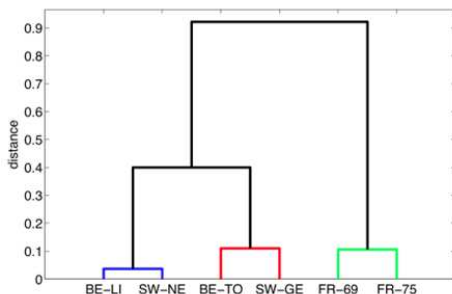


FIGURE 2 – Classification émergente pour le paramètre de la vitesse d'articulation.

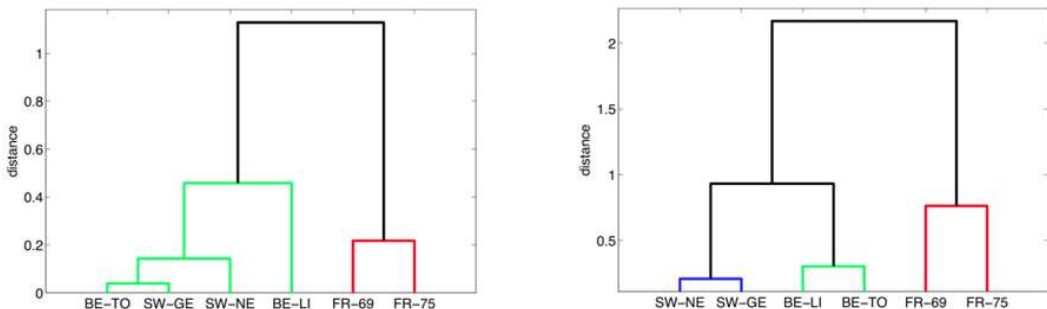


FIGURE 3 – Classifications émergentes pour le paramètre du débit (gauche) et de la pente de Fo (à droite).

3.3 Discussion

Selon que l'on choisisse une approche guidée par les hypothèses ou une approche émergente, les 8 paramètres prosodiques mesurés ne présentent pas tous la même efficacité discriminatoire. Le tableau 3 résume les mesures qui se révèlent discriminantes dans les deux approches (++), celles qui ne le sont que dans une seule approche (+) et celle qui ne l'est dans aucune (-) :

Hyp. I: mesures de rythme et de débit	
Vitesse d'articulation	++
Débit	+
Densité accentuelle	+
Poids métrique du SA	-
ΔC (x100)	++
%V	-
Hyp. II: mesures d'accentuation	
lengthening	-
FO rise	+

TABLE 3 – Fiabilité des mesures prosodiques calculées pour la discrimination prosodique des variations régionales dans les deux approches (approche guidée par les hypothèses vs approche émergente).

Deux caractéristiques conduisent à la classification attendue avec des différences significatives en regard de H1 dans les deux conditions : la vitesse d'articulation et la mesure de ΔC . Le débit et la densité accentuelle correspondent à la classification prévue dans une approche guidée par les hypothèses uniquement, le poids métrique du SA permet d'aboutir à une classification cohérente mais n'est significative dans aucune des deux situations. Quant au paramètre %V, il n'est significatif dans aucune des deux conditions. Aucune des mesures accentuelles choisies ne confirme l'hypothèse H2, quelle que soit l'approche : la pente de FO n'est pas significative dans une approche émergente, et l'allongement pénultième aboutit à des classifications cohérentes mais sans différence significative. Cette absence de pouvoir discriminant s'explique aisément quand on sait que seuls certains SA sont marquées d'une proéminence dans les variétés de français parlées en Belgique et en Suisse. Ainsi, la moyenne statistique des caractéristiques d'un locuteur peut masquer des différences importantes, puisqu'elles ne sont pas systématiques. Bien qu'il soit difficile de comparer nos résultats avec des études qui n'ont pas été menées avec le même matériel et la même méthodologie, nos résultats confirment que les locuteurs de Suisse et de Belgique ont tendance à parler plus lentement que les locuteurs de variétés de français parlées dans le domaine d'oïl, et notre étude montre qu'un tel état de fait est principalement dû à deux indices différents : la vitesse d'articulation et ΔC (ce dernier indiquant que les variétés d'oïl ont une structure syllabique plus régulière que les autres variétés de français).

4 Conclusion

Dans cet article, nous avons cherché à voir si l'on pouvait classer plusieurs variétés de français sur une échelle de d'éloignement dialectal par rapport à des critères uniquement prosodiques. Des mesures statistiques robustes ont été proposées en vue d'estimer les caractéristiques des variétés considérées, et les méthodes de classification non-supervisée

ont été introduites pour regrouper les variétés par rapport à leurs caractéristiques moyennes. Une classification guidée par les hypothèses permet d'aboutir au classement prédit, alors qu'une approche émergente débouche sur une configuration plus contrastée. À l'avenir, de nouvelles études seront nécessaires afin d'évaluer la pertinence des différences prosodiques entre les variétés dialectales étudiées ici (alignement tonal, prise en compte d'autres contextes de relativisation pour juger du caractère +/- proéminent de la pénultième, traitement de données conversationnelles, etc.), et des tests de perception devront permettre de valider l'échelle d'éloignement dialectal sur laquelle nous avons classé les variétés examinées dans cet article.

Références

AVANZI, M., SIMON, A. C., GOLDMAN, J.-P., AUCHLIN, A. (2010). "C-PROM: An Annotated Corpus for French Prominence Study". *Proceedings of Prosodic Prominence, Speech Prosody 2010 Workshop*, Chicago.

BOULA DE MAREÛIL, P., BARDIAUX, A., (2011). "Perception of French, Belgian and Swiss accents by French and Belgian listeners". *Proceedings of the 4th ISCA Tutorial and Research Workshop on Experimental Linguistics*, 47–50.

BOERSMA, P., WEENINK, D. (2012). "Praat, version 5.5", www.praat.org.

CARTON, F., ESPESSER, R., VAISSIÈRE, J. (1991). "Etude sur la perception de l'accent régional du Nord et de l'Est de la France", *Proc. 12th ICPHS*, Aix-en-Provence, 422-425.

DURAND, J., LAKS, B., LYCHE, C. (2009). *Phonologie, variation et accents du français*. Paris : Hermès.

GOLDMAN, J.-P. (2011). "EasyAlign: an automatic phonetic alignment tool under Praat". *Proceedings Interspeech 2011*, 3233-3236.

JUN, S. A., FOUGERON, C. (2002). "Realizations of Accentual Phrase in French intonation". *Probud*, 14, 147-172.

MAHMOUDIAN, M., JOLIVET, R. (1984). "L'accent vaudois". In *Encyclopédie illustrée du Pays de Vaud*, Éditions 24Heures.

RAMUS, F., NESPOR, M., MEHLER, J. (1999). "Correlates of linguistic rhythm in the speech signal", *Cognition*, 73/3, 265-292

SCHWAB, S., RACINE, I. (à par.). "Le débit lent des Suisses romands : mythe ou réalité? "

SCHWAB, S., AVANZI, M., GOLDMAN, J.-P., MONTCHAUD, P., RACINE, I. (2012). "An Acoustic Study of Penultimate Accentuation in Three Varieties of French", *Proceedings of Speech Prosody 2012*, Shanghai.

STERLING-MILLER, J. (2007). *Swiss French Prosody: Intonation, Rate, and Speaking Style in the Vaud Canton*, PhD, Illinois University.

TREVOR, H., TIBSHIRANI, R., FRIEDMAN, J. (2009). *Hierarchical clustering. The Elements of Statistical Learning*, New York, Springer.