

**DÉVELOPPEMENT D'UN MODÈLE PHONÉTIQUE MINIMAL DE
L'ACCENT QUÉBÉCOIS. EFFET DE L'INTRODUCTION DE
TRAITS PHONÉTIQUES CARACTÉRISTIQUES SUR
L'IDENTIFICATION DE LA PROVENANCE D'UN LOCUTEUR**

Claude Paradis et Jean Dolbec

C.I.R.A.L., Université Laval, Québec

Claude.Paradis@fl.ulaval.ca

Université du Québec à Chicoutimi, Chicoutimi

Jdolbec@uqac.quebec.ca

Résumé: Les résultats d'une expérience de perception montrent que des sujets québécois sont capables d'identifier la provenance d'un sujet francophone à partir de la présence de certains marqueurs phonétiques. La facilité de cette identification est cependant fonction du nombre et de la nature des variantes présentes dans l'énoncé.

Mots clés: phonétique, variation, accent, français, Québec, sociolinguistique

1. INTRODUCTION

Le présent travail s'inscrit dans le cadre d'un programme de recherche plus vaste visant l'intégration de la variation socio-géographique dans la parole de synthèse en vue d'en augmenter le naturel et d'en favoriser l'acceptation par les utilisateurs. Dans des recherches antérieures, nous avons essayé de démontrer la faisabilité d'une modélisation de la variation phonétique existant en français du Québec et d'Acadie (Dolbec et Paradis, 1991). Nous avons aussi tenté de mesurer, par une série de tests de perception et d'analyses, l'effet de la variété linguistique sur l'intelligibilité de mots isolés (Paradis *et al.*, 1993).

2. PROBLEMATIQUE ET OBJECTIF DE LA RECHERCHE

À mesure que la technologie s'améliore et que nos connaissances sur les conditions d'usage des variétés linguistiques constituant le répertoire linguistique d'une communauté se développent, les chercheurs dans le domaine de la synthèse de la parole tentent d'intégrer dans leurs systèmes une partie de la variation sociale et stylistique propre à une langue, qu'il s'agisse du nombre de voix disponibles ou des divers accents possibles pour une langue donnée (Fant, 1991; Karlsson, 1991; Carlson, 1992).

Dans le but de minimiser les coûts d'une adaptation éventuelle d'un système de synthèse en français européen à une variété de français québécois, nous avons entrepris d'identifier les traits perçus comme étant les plus caractéristiques du phonétisme du français québécois ainsi que les combinaisons minimales de traits ayant un effet sur l'identification d'un locuteur comme originaire du Québec. On peut en effet s'interroger, lorsqu'on identifie un "accent" quelconque, sur les processus mis en oeuvre et sur les éléments pris en compte lors de cette identification. Certains éléments fonctionnent certainement comme des schibboleths alors que d'autres doivent sans doute revenir à une certaine fréquence ou être mis en rapport avec d'autres pour qu'un "accent" soit détecté avec certitude. Ainsi, en contexte québécois, si l'utilisation de l'expression *fais pas simple* trahit les personnes originaires du Saguenay et le *r roulé* celles de l'ouest de Trois-Rivières, une seule occurrence d'un *ch* affaibli ne peut pas permettre une identification incontestable d'un Beauceron. D'autres éléments doivent être pris en compte. La mise au point de ce modèle phonétique minimal de l'accent québécois repose sur la cueillette et l'analyse de jugements portés par des sujets québécois dans le cadre de tests de perception.

3. MÉTHODOLOGIE

3.1. *Test de perception*

Le premier test de perception que nous avons conduit, et dont nous présentons ici les résultats, repose sur l'hypothèse qu'un sujet ayant le français québécois comme langue maternelle est en mesure d'identifier, au moins sommairement, la provenance d'un francophone si un certain nombre de traits caractéristiques d'une variété de français ou d'une autre sont présents dans son discours. Ce test est construit de façon à permettre d'identifier les traits phonétiques ou les combinaisons de traits phonétiques qui ont un effet sur la reconnaissance de la provenance d'un sujet parlant français.

3.2. *Choix des séquences*

Afin de contrôler les effets liés à la longueur des phrases, à la prosodie et au nombre de variantes phonétiques, les séquences testées sont toutes des phrases de type déterminant+N+de+(déterminant)+N+est+[ADJ. ou PART. PASSÉ] (ex. *La sortie de la pente est rapide*), réalisées prosodiquement en trois groupes accentuels. En syllabe accentuée, se retrouvent des configurations pouvant donner lieu à des prononciations typiques du français du Québec ou du français de France, ainsi que quelques configurations neutres, c'est-à-dire ne pouvant être associées ni à l'une ni à l'autre variété de français. Les traits caractéristiques du français québécois retenus sont l'affrication de t/d, l'ouverture de la voyelle fermée /i/ en syllabe

entravée et la réalisation antérieure et plus termée de la nasale (an). Se retrouvent aussi en fin de syntagme accentuel ou tonal, les mots *agités*, *Paul*, *abeille*, *échecs*, et *terminée* qui ne donnent pas lieu à des prononciations significativement différentes en français québécois et en français de France. Ce sont les variantes neutres.

Sur les 343 séquences à trois syntagmes accentuels possibles, seulement quarante-huit (48), représentatives de diverses combinaisons, ont été testées. Elles ont été choisies pour représenter 5 types de séquences: *Homogènes I* (trois fois la même variante), *Homogènes II* (deux variantes identiques plus une autre variante de la même variété), *Homogènes III* (trois variantes différentes de la même variété); *Homogènes + Neutres* (une ou deux variantes de la même variété plus une ou deux variantes neutres); *Conflictuelles* (variantes de l'une et l'autre variétés).

3.3. Enregistrement des stimuli

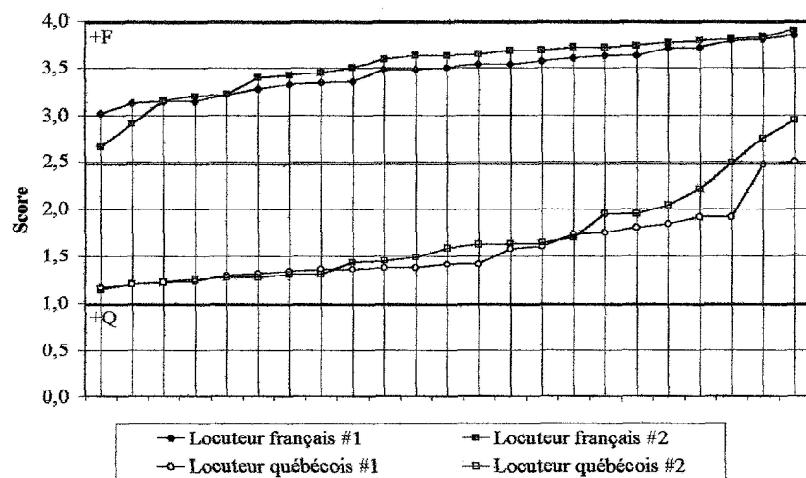
On a demandé à un sujet bilingue, c'est-à-dire à un sujet capable de produire aussi bien les variantes québécoises que les variantes françaises des variables retenues, de lire les séquences élaborées. Une attention toute particulière était portée à une réalisation uniforme de la prosodie. Afin de pouvoir introduire des « distracteurs », on a demandé également à quatre sujets « unilinguistes » de lire les phrases retenues dans leur variété naturelle de français soigné. Deux des quatre sujets possédaient un répertoire de variétés de français européen alors que deux autres avaient comme répertoire vernaculaire un ensemble de variétés de français québécois. Les séquences ont été produites en studio et enregistrées sur du matériel professionnel.

3.4. Sujets

Les séquences ont ensuite été soumises à 56 étudiants de premier cycle à l'Université Laval (Québec) n'ayant jamais suivi de cours de phonétique. Après avoir entendu chacune des séquences du test, les sujets devaient identifier, avec plus ou moins de certitude, l'origine géographique du locuteur. Ils avaient le choix entre quatre réponses: 1. *Je suis certain(e) que c'est un Québécois*; 2. *Je pense que c'est un Québécois*; 3. *Je pense que c'est un Français*; 4. *Je suis certain(e) que c'est un Français*.

3. RÉSULTATS

Avant d'examiner les résultats pour le sujet bilingue utilisé, jetons un coup d'œil sur les résultats obtenus pour les quatre sujets dont les séquences ont été utilisées pour « masquer » celles du sujet bilingue. À l'examen de la Figure 1, laquelle présente en ordre croissant la moyenne des réponses obtenues pour chacun des énoncés, on constate que les locuteurs ayant un répertoire de variétés de français européen sont identifiés comme venant de France et ceux ayant un répertoire de variétés de français québécois sont perçus comme originaires du Québec. Ces quatre courbes montrent donc que des sujets québécois sont capables d'identifier, au moins grossièrement, l'origine géographique d'un francophone sur la base de son phonétisme.



Le Tableau 1 permet de mieux apprécier l'effet de différentes combinaisons de variantes sur cette identification. Un résultat s'écartant de zéro dans un sens ou dans l'autre indique un plus grand degré de certitude dans l'identification comme québécois (valeurs négatives) ou comme français (valeurs positives). Pour les variantes québécoises, le degré de certitude augmente régulièrement de *Homogène I* à *Homogène III*, soit avec le nombre de variantes différentes présentes; le résultat est moins net pour les variantes françaises même si le meilleur résultat est encore une fois obtenu avec les séquences *Homogènes III*. Dans un cas comme dans l'autre, l'introduction de variantes neutres ou conflictuelles, réduit de façon très importante le degré de certitude dans la reconnaissance.

Tableau 1 : Force de l'identification de la provenance par types de séquences et par variété

Types de séquences	Q	F
Homogènes I	-1,18	+0,96
Homogènes II	-1,33	+0,74
Homogènes III	-1,40	+1,07
Homogènes + Neutres	-0,85	+0,39
Conflictuelles	+0,06	-0,35

4. CONCLUSION

À première vue, les résultats obtenus semblent infirmer l'hypothèse minimaliste puisque la quantité et la diversité des informations présentes semblent avoir un effet sur le taux de reconnaissance de l'origine géographique, les meilleurs résultats étant obtenus lorsqu'on a trois variantes différentes d'une même variété (*Homogènes III*). Mais il convient toutefois de nuancer cette affirmation. En effet, compte tenu que les variantes prises en compte constituent des caractéristiques capitales du français québécois, il n'est pas dit que ces quelques variantes ne permettraient pas de faire l'économie d'une multiplicité de variantes secondaires.

RÉFÉRENCES

- Carlson, R. (1992). Synthesis: modelling variability and constraints, In : *Speech Communication*, Vol. 11:2-3, pp. 159-166.
- Dolbec, J. et C. Paradis (1991). Modélisation de la variation phonétique en français québécois, In : *Revue québécoise de linguistique théorique et appliquée*, Vol. 10:3, pp. 85-106.

- Fant, G. (1991). What can basic research contribute to speech synthesis?, In : *Journal of Phonetics*, **Vol. 19:1**, pp. 75-90.
- Karlsson, I. (1991). Female voice in speech synthesis, In : *Journal of Phonetics*, **Vol. 19:1**, pp. 111-120.
- Paradis, C., Brousseau, M. et J. Dolbec (1993). Variétés linguistiques et intelligibilité: enjeux sociolinguistiques pour la synthèse de la parole. In : *Revue québécoise de linguistique*, **Vol. 22:2**, pp.13-36.