Document generated on 07/19/2025 1:03 a.m.

Revue québécoise de linguistique

Variétés linguistiques et intelligibilité : enjeux sociolinguistiques pour la synthèse de parole

Claude Paradis, Martin Brousseau and Jean Dolbec

Volume 22, Number 2, 1993

Sociolinguistique

URI: https://id.erudit.org/iderudit/602768ar
DOI: https://doi.org/10.7202/602768ar

See table of contents

Publisher(s)

Université du Québec à Montréal

ISSN

0710-0167 (print)
1705-4591 (digital)

Explore this journal

Cite this article

Paradis, C., Brousseau, M. & Dolbec, J. (1993). Variétés linguistiques et intelligibilité : enjeux sociolinguistiques pour la synthèse de parole. *Revue québécoise de linguistique*, *22*(2), 13–36. https://doi.org/10.7202/602768ar

Article abstract

This paper is related to the opportunity of using a vernacular variety in speech synthesis instead of a more formal one. The results of three experiments, especially those of the more elaborated one based on a lexical decision task, point out that stimuli spoken in a Quebec French variety are more easily processed by Quebec French speaking subjects than those spoken in an European variety. These results suggest that the linguistic variety acts as a *filter* that may facilitate or, on the contrary, impede and slow down the processing of words.

Tous droits réservés © Université du Québec à Montréal, 1993

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

Revue québécoise de linguistique, vol. 22, n° 2, 1993, © RQL (UQAM), Montréal
Reproduction interdite sans autorisation de l'éditeur

VARIÉTÉS LINGUISTIQUES ET INTELLIGIBILITÉ: ENJEUX SOCIOLINGUISTIQUES POUR LA SYNTHÈSE DE PAROLE*

Claude Paradis
Martin Brousseau
Jean Dolbec

1. Introduction

Le rapprochement entre sociolinguistique et synthèse de la parole peut paraître surprenant à première vue; il prend cependant rapidement sens lorsque l'on s'avise que l'objectif ultime de la recherche en synthèse comme en reconnaissance de la parole est la mise au point de systèmes dont la performance s'approcherait de celle d'un locuteur humain. Pour ce faire, les développements strictement technologiques doivent aller de pair avec l'incorporation du maximum d'informations linguistiques, non seulement de nature phonétique ou phonologique, mais aussi syntaxique, sémantique ou pragmatique. À ce titre, la variation a évidemment sa place en tant que composante essentielle du fonctionnement du système linguistique.

Nous travaillons depuis un certain temps à évaluer l'intérêt, en contexte québécois, de la prise en compte de la variation sociolectale dans des applications en synthèse de la parole, cf. Dolbec & Paradis (1990, 1991). Posant comme hypothèse que l'utilisation d'une variété linguistique vernaculaire pouvait contribuer à améliorer l'intelligibilité, nous avons mené un certain nombre d'expériences mettant en cause des mots isolés; les résultats de deux d'entre elles ont été présentés dans Dolbec, Paradis, Brousseau & Deshaies (1991). Après un bref rappel des enjeux de

* La réalisation de cette recherche a été rendue possible grâce à une subvention du Fonds pour la formation des chercheurs et l'aide à la recherche du Québec (FCAR 92-ER-1111).

notre recherche et des premiers résultats obtenus, on trouvera dans les pages qui suivent la description d'une troisième expérience, impliquant une tâche de décision lexicale, qui se distingue des deux premières tant par les raffinements importants apportés au protocole expérimental que par la netteté des résultats.

2. Problématique

Si le problème de la variabilité linguistique sous ses diverses formes (coarticulatoire, linguistique, individuelle ou sociogéographique) s'est trouvé très tôt au coeur des préoccupations en reconnaissance de la parole où il constitue toujours un défi majeur pour les systèmes multi-locuteurs en parole continue, cf. Lee (1989), ce n'est au contraire que tout récemment qu'il a commencé à se manifester dans le champ des recherches en synthèse de la parole. Cette émergence tardive, qui n'est pas sans ressemblance avec la situation de la sociolinguistique dans l'histoire de la linguistique, peut s'expliquer de diverses façons. D'un point de vue strictement technologique, il est évident que de tels raffinements ne deviennent envisageables qu'à partir du moment où les systèmes ont atteint un degré de flexibilité suffisant pour ce faire, ce qu'on commence tout juste à entrevoir. Par ailleurs, les applications pour les premiers systèmes rudimentaires impliquaient le plus souvent des situations fortement contraintes permettant l'utilisation de sous-ensembles restreints (chiffres, vocabulaire restreint de mots) et reposaient sur une adaptation de l'utilisateur à la machine plutôt que l'inverse. Il ne faudrait pas négliger non plus le poids des facteurs économiques voire idéologiques; non seulement le coût de développement des systèmes n'a pas favorisé leur multiplication en fonction de marchés fragmentés, mais encore la composition des équipes de chercheurs, leur localisation et leurs sources de financement ont tendu à privilégier la variété de prestige au détriment des variétés plus périphériques.

La situation est en train de changer avec l'apparition dans toutes les sphères de la vie courante d'un nombre sans cesse croissant d'applications utilisant la parole de synthèse, en même temps que le progrès technologique dans le développement des systèmes laisse entrevoir de nouvelles possibilités. Aussi ne faut-il pas s'étonner que des voix de plus en plus nombreuses se fassent entendre pour appeler au développement de systèmes moins monolithiques et plus flexibles, qu'il s'agisse du type de voix disponibles, en particulier en fonction du sexe et de l'âge, cf. Fant (1991); Karlsson (1991), de l'intégration des marques d'émotivité, cf. Fant (1991), ou de la prise en compte des différents styles, registres ou accents, cf. Bladon & al.

(1987), et même des caractéristiques individuelles du locuteur, cf. Carlson (1992). Ces différents aspects qui visent tous à améliorer l'intelligibilité et le naturel des systèmes de synthèse, et en définitive à en favoriser l'acceptabilité dans une multitude de situations dans la mesure où leur output se rapproche encore plus du langage humain, mettent tous en cause sous une forme ou l'autre la variabilité linguistique. D'où la conclusion d'un article récent de Carlson (1992, p. 164) à l'effet que «*speaker variation* and *speaker variability* are keywords in future synthesis research.»

En tant que sociolinguistes et phonéticiens travaillant au Québec, nous nous intéressons à l'intégration de la variation phonétique dans des systèmes de synthèse de parole en français; nous avons d'ailleurs mis au point un ensemble de routines de modélisation de la variation phonétique en français québécois qui pourraient s'intégrer à un système de synthèse pour permettre l'obtention de réalisations phonétiques plus diversifiées et plus conformes à l'usage réel de différents types de locuteurs caractérisables par leur profil sociogéographique, cf. Dolbec & Paradis (1991); Paradis & Dolbec (1992).

S'il peut, en théorie, paraître souhaitable d'augmenter le naturel des systèmes de synthèse en y incorporant la possibilité de produire diverses variétés linguistiques en fonction du message et du récepteur en cause, il y a tout de même lieu de se demander, compte tenu des coûts impliqués, si l'ajout d'une telle composante variationniste est rentable, c'est-à-dire si elle contribue de façon significative à la performance globale du système, ou s'il ne s'agit que d'un raffinement relativement accessoire. Divers paramètres peuvent être pris en considération dans l'évaluation des systèmes de synthèse, les uns objectifs comme l'intelligibilité du message, les autres plus subjectifs et faisant intervenir des jugements sur le naturel, le confort d'écoute et l'acceptabilité globale.

Le caractère particulièrement complexe de la question des attitudes subjectives des locuteurs québécois envers les différentes variétés linguistiques a été signalé dans Dolbec & Paradis (1990). Les travaux de Méar-Crine & Leclerc (1976), de Lappin (1982) et de Tremblay (1990) sur le français québécois montréalais permettent en effet de croire que, si les variétés standard sont généralement mieux évaluées en situation de test ou d'enquête que les variétés non standard ou familières, ces dernières n'en font pas moins l'objet d'évaluations plus positives pour certains groupes de locuteurs, sous certains aspects et dans certaines situations. Les résultats de Lappin (1982) vont plus loin en faisant voir l'existence d'un clivage entre des

caractéristiques québécoises bien acceptées et d'autres encore fortement stigmatisées, ce qui accrédite l'hypothèse d'une certaine évolution de la norme. Il faudrait en outre tenir compte de l'affirmation récente d'une identité québécoise qui entraîne souvent le rejet d'une norme sentie comme étrangère, comme cela pourrait être le cas pour la variété utilisée dans les systèmes de synthèse de conception européenne. Il semble donc que l'on ait affaire à une interaction subtile de deux normes de prestige, cf. Labov (1976, pp. 250-251) et Trudgill (1974, pp. 96-101), l'une explicitement reconnue (*overt prestige*) et l'autre plus ou moins occulte (*covert prestige*), interaction dont il apparaît bien difficile de cerner toutes les facettes, mais qui joue de toute façon en faveur de la conception de systèmes de synthèse pouvant intégrer la variation. Dans une perspective de synthèse de la parole, la question des attitudes face à la variété linguistique a été abordée indirectement par Robert, Choinière & Descout (1989, p. 642) qui tendent cependant à minimiser l'importance de ce facteur:

> The three systems have a parisian accent [...] according to French-speaking Quebecois subjects. This accent does not seem to be embarrassing for the perception of voice quality since Multivox is the best system according to the subjects while it has the most parisian accent [...]. The parisian accent is not a barrier for naturalness and acceptability of synthetic voices heard by Quebecois.

Il convient toutefois d'attirer l'attention sur le fait qu'aucun des trois systèmes n'utilisait une variété québécoise et que leurs performances intrinsèques étaient par ailleurs très inégales, ce qui n'a sans doute pas manqué d'avoir une influence sur les jugements portés.

La recherche dont il est fait ici état s'intéresse directement pour sa part à la dimension centrale de l'intelligibilité qui constitue probablement le déterminant principal de la performance globale d'un système de synthèse, voir Pavlovic, Rossi & Espesser (1990). Il apparaît en effet essentiel, si l'on veut promouvoir le développement de systèmes pouvant prendre en compte la variation linguistique, et plus spécifiquement la variation sociophonétique, d'évaluer dans quelle mesure l'utilisation d'une variété linguistique donnée a un effet sur le traitement du message linguistique véhiculé, c'est-à-dire sur la capacité, la vitesse et la précision du décodage par le sujet récepteur. Nous pouvons, d'entrée de jeu, poser comme hypothèse que plus la variété linguistique utilisée est familière pour un sujet ou un ensemble de sujets, plus l'opération de décodage est susceptible de s'effectuer rapidement et efficacement. S'il existe, dans le cadre des recherches sur le bilinguisme ou le

plurilinguisme, un certain nombre de travaux sur le traitement mental des messages livrés dans des langues différentes, nous ne connaissons par contre aucune étude empirique qui porte spécifiquement sur la facilité de décodage ou de traitement de variétés linguistiques associées à des groupes sociaux ou à des *styles* à l'intérieur d'une seule et même langue.

3. Expériences antérieures

Tel que mentionné précédemment, deux expériences, l'une d'identification de stimuli verbaux entendus, l'autre d'appariement d'un stimulus sonore avec un stimulus visuel, ont été menées antérieurement. Il nous semble utile, à des fins de comparaison et pour permettre de mieux apprécier la portée des résultats présentés ici, d'en rappeler brièvement l'essentiel[1].

3.1 Expérience I

La première expérience visait à évaluer la capacité d'un certain nombre de sujets à reconnaître des mots produits en français québécois et en français européen. Vingt mots présentant des réalisations phonétiques caractéristiques du français québécois[2] ont été produits par un locuteur *bilectal*[3] dans une variété de français québécois (VQ) et dans une variété de français européen (VE)[4] pour un total de 40 stimuli, lesquels ont ensuite été présentés à 60 sujets ayant le français québécois comme langue maternelle. Les sujets avaient comme consigne de transcrire sur une feuille-réponse le mot entendu. Les résultats présentés à la figure 1 montrent que les sujets ont globalement mieux réussi à identifier les stimuli en VQ que ceux en VE. En effet, alors que les erreurs d'identification sont au nombre de 108 pour les

---

[1] On se référera pour une présentation détaillée à Dolbec & al. (1991).

[2] Les phénomènes phonétiques caractéristiques retenus pour cette expérience et la suivante sont les mêmes que ceux de l'expérience III; voir ci-dessous, notamment le tableau 1.

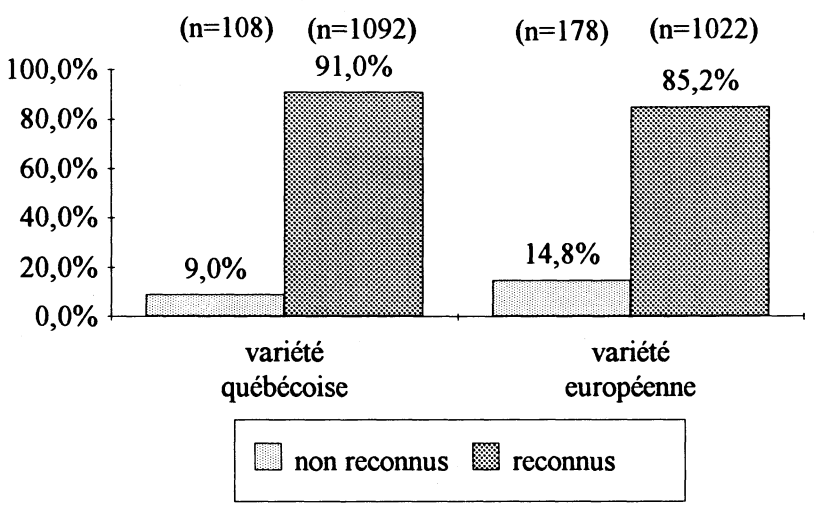
[3] Un locuteur bilectal est un locuteur qui possède quasi parfaitement deux variétés linguistiques, notamment, dans le cas qui nous concerne, une variété de français québécois et une variété de français européen. Le locuteur qui a enregistré les stimuli pour toutes les expériences rapportées ici est né au Québec, vit actuellement au Québec, mais a fait de longs séjours en France où il a des parents, est très doué pour l'apprentissage des langues et, enfin, a un talent hors du commun pour imiter divers accents.

[4] Nous entendons par variété de français québécois une variété de français qui a été unanimement reconnue comme telle sur la base de ses caractéristiques phonétiques par trois professeurs et quatre étudiants de phonétique et, inversement, par variété de français européen, une variété de français que les mêmes évaluateurs ont identifiée comme n'étant pas du français québécois et caractérisée globalement comme du français de France, sans égard aux diverses sous-variétés qu'il conviendrait de distinguer dans chaque cas.

stimuli en VQ (9 %), elles s'élèvent à 178 pour les stimuli produits en VE (15%), la différence étant significative à ,001 (χ^2=18,89, d.l.=1)[5].

Même si les résultats de cette expérience qui font voir un avantage significatif pour la variété québécoise sont intéressants, ils ne peuvent cependant pas être considérés comme décisifs dans la mesure où le protocole relativement simple utilisé permettait difficilement de contrôler les interférences de facteurs autres que strictements phonétiques (conditions d'écoute, ordre de présentation des stimuli, effets d'apprentissage, familiarité avec les mots, associations sémantiques des mots, etc.). Ne prenant en considération que la justesse de la réponse sans égard au facteur temps de réaction, il ne fournissait par ailleurs aucune indication sur la plus ou moins grande difficulté de traitement selon la variété linguistique, toutes les bonnes réponses comme toutes les erreurs étant considérées comme équivalentes de ce point de vue. Il devient alors difficile d'apprécier dans quelle mesure cette différence dans le taux d'identification (91% pour VQ contre 85% pour VE) aurait une incidence effective dans le cadre d'une phrase ou d'un texte suivi où le contexte exerce un effet désambiguïsant, voire prédictif, important.

Figure 1
Expérience I: pourcentages d'identification en fonction de la variété linguistique

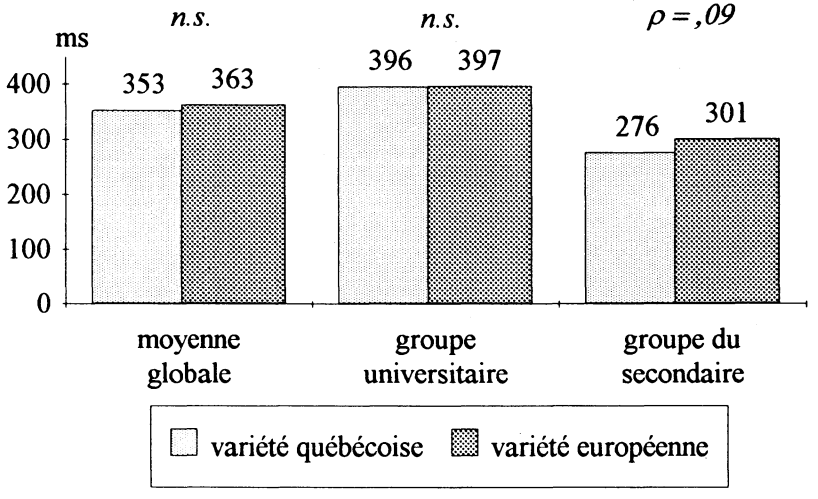


[5] Sauf mention contraire, le test statistique utilisé dans cet article est celui du chi-deux.

3.2 Expérience II

Compte tenu de ces premiers résultats, la deuxième expérience avait pour objectif de vérifier si l'utilisation d'une variété familière pouvait avoir un effet facilitant sur la perception et l'intégration d'un stimulus, ce qui implique une moins grande mobilisation des ressources cognitives qui peuvent alors être rendues disponibles pour d'autres tâches (traitement syntaxique, interprétation sémantique, etc.).

Figure 2
Expérience II: moyennes des temps de réaction en fonction de la variété linguistique et du groupe de sujets



Le protocole mis au point visait à déterminer si un stimulus produit dans une variété linguistique familière était plus rapidement perçu et traité que le même stimulus produit dans une variété non familière. Les 40 stimuli de l'expérience I ont été réutilisés. La tâche consistait, pour les sujets, à associer le plus rapidement possible un stimulus sonore produit soit en VQ, soit en VE, avec l'un des deux mots affichés sur un écran cathodique. Les sujets répondaient en appuyant sur une touche située à gauche ou à droite d'un clavier de micro-ordinateur selon que le stimulus oral correspondait au stimulus visuel affiché à gauche ou à droite de l'écran. C'est le temps écoulé entre la fin de la production du stimulus oral et la

réponse du sujet, mesuré en millisecondes, qui constituait la variable dépendante. Le lieu d'affichage des stimuli de l'une et de l'autre variété ainsi que leur ordre d'apparition ont été contrôlés dans le protocole utilisé. Deux groupes de sujets, 38 étudiants universitaires et 38 jeunes adultes inscrits à l'éducation des adultes (ordre secondaire), tous droitiers, ont été soumis au test de perception. Pour le premier groupe de sujets universitaires, aucune différence n'a été constatée dans les temps de réaction moyens à l'une ou l'autre variété alors qu'on a observé une différence marginalement significative (ρ =,09) de 17 ms en faveur de la variété québécoise dans le groupe des sujets moins scolarisés (figure 2).

3.3 Enseignements tirés des expériences I et II

Si les résultats de l'expérience I appuient l'hypothèse que les mots produits dans une variété familière de français sont plus facilement identifiés, ceux de l'expérience II n'autorisent pas à conclure qu'il en résulterait une économie dans le processus de traitement. Divers facteurs parasites peuvent être invoqués pour expliquer ces résultats discordants. Il semble en effet que certains stimuli en variété québécoise aient entraîné des problèmes de décodage au moins aussi importants que les stimuli en variété européenne, vraisemblablement parce qu'ils comportaient des réalisations extrêmes[6] de certains allophones. Par ailleurs, des lacunes dans le protocole de l'expérience II semblent avoir favorisé l'utilisation de stratégies de réponse qui ont pu fausser les résultats[7]. C'est probablement le recours à de telles stratégies de simplification, combiné avec une trop grande facilité de la tâche demandée, qui explique le nombre anormalement élevé (47%) de réponses anticipatives chez les sujets d'éducation secondaire qui paraissent avoir privilégié l'élément vitesse au détriment de la précision.

---

[6] Il pouvait s'agir soit de variantes très stigmatisées, soit de réalisations moins familières aux sujets testés, comme c'est le cas pour certaines formes de diphtongaison plus courantes à Montréal que dans la région de Québec.

[7] Certains sujets ont avoué en fin de test qu'ils ne se concentraient que sur un des deux stimuli visuels fournis (le plus souvent celui de gauche); ils se trouvaient ainsi à redéfinir la tâche puisqu'ils se contentaient de décider si le mot entendu correspondait ou non à ce mot privilégié. Le recours à une telle stratégie serait par ailleurs à mettre en relation avec le fait que les temps de réaction sont significativement plus faibles (347 ms contre 363 ms; $\rho < ,05$) lorsque le mot cible occupait la position gauche de l'écran plutôt que la droite.

4. Nouvelle expérience et discussion

4.1 Objectifs

La troisième expérience visait expressément à corriger une partie des problèmes rencontrés en particulier quant à l'homogénéité des stimuli et à la trop grande facilité de la tâche. Elle s'inspire des résultats obtenus par Meyer, Schvaneveldt & Ruddy (1975) dans une série de tests de perception visuelle. Ces derniers ont en effet montré que le temps de réaction de même que le taux d'erreur dans une tâche de décision lexicale qui implique des paires de mots[8] se trouvent significativement réduits lorsque le second mot de la paire est sémantiquement associé au premier comme c'est le cas dans BUTTER-BREAD par opposition à NURSE-BREAD. Ces résultats suggèrent à Meyer & al. (1975, p. 102) deux hypothèses qui permettent de rendre compte de cet effet facilitant du contexte sémantique:

> The first assumption is that words are stored at distinct «locations» in lexical memory and the memory is organized semantically, so that in some sense, associated words like BREAD and BUTTER are relatively close together whereas unassociated words like NURSE and BUTTER are farther apart. The second assumption is that accessing information from a given memory location produces residual neural activity that spreads to other nearby locations. This temporary increase of excitation then produces the faster recognition of associated words.

Notre objectif ne consiste pas ici à vérifier si le modèle de propagation-excitation de Meyer (*spreading-excitation model*) peut également s'appliquer à la perception auditive, mais plutôt à chercher à voir, en transposant cet effet facilitant du contexte du plan sémantique au plan phonétique, si la variété linguistique utilisée pour un premier stimulus peut agir de façon similaire, en activant un *filtre*[9]

---

[8] La procédure originellement utilisée par Meyer & al. (1975) impliquait des chaînes de caractères associées par paires. Une pause de 500 ms séparait la présentation de chaque paire et les deux stimuli d'une même paire étaient affichés l'un à la suite de l'autre avec un intervalle de 500 ms. La tâche des sujets consistait à juger dans chaque cas si la chaîne de caractères constituait un mot ou non (un logatome). Les sujets répondaient en pressant l'une des deux touches correspondant aux réponses possibles, à savoir *oui* ou *non*. Le temps de réponse était mesuré entre l'apparition de la chaîne de caractères et le moment où le sujet enfonçait une touche.

[9] En psychologie, le concept de *filtre* a été développé par Broadbent (1958) et s'inscrit dans la description d'un modèle d'attention sélective. Bien que le modèle proposé par Broadbent ait souvent été remis en cause, la grande majorité des chercheurs qui s'intéressent à la question admettent qu'il existe, dans tout processus de perception, des modes de sélection de l'information. Ce consensus nous

qui facilitera ou, au contraire, gênera la perception du deuxième stimulus. Nous faisons l'hypothèse que l'effet de filtre se traduira par une diminution ou une augmentation du taux d'erreur et/ou du temps global de réaction pour la décision lexicale selon que le filtre mis en place sera familier ou non d'une part, et qu'il y aura concordance ou non entre la variété linguistique du stimulus et le filtre déjà mis en place d'autre part.

4.2 Méthodologie

La procédure expérimentale utilisée pour ce test se présente comme une adaptation à nos besoins de celle mise au point par Meyer & al. (1975).

4.2.1 Stimuli

Les stimuli pour cette troisième expérience consistent en quarante-huit (48) monosyllabes représentatifs de douze (12) phénomènes phonétiques considérés comme typiques des variétés de français québécois et, inversement, comme non présents dans la variété européenne; il s'agit dans tous les cas de phénomènes partagés par l'ensemble de la communauté et qui ne font pas l'objet d'évaluations négatives. Comme il se doit pour une tâche de décision lexicale, on retrouve le même nombre de mots (24) que de logatomes (24), ces derniers étant construits de façon à former une paire minimale avec le mot auquel ils se trouvent associés. La liste des stimuli est présentée au tableau 1.

Chaque mot et chaque logatome ont été prononcés par un seul locuteur *bilectal* dans les deux variétés sociogéographiques de français testées, à savoir VQ et VE, pour un total de 96 stimuli[10], lesquels ont d'abord été enregistrés sur bande magnétique pour être ensuite numérisés à 20 kHz à l'aide du logiciel de traitement du signal acoustique *CSL* de Kay Elemetrics Corp. Six jeux de stimuli contenant 48 paires chacun ont ensuite été créés de façon à équilibrer la distribution des stimuli et

---

suffit pour justifier la formulation de notre hypothèse et chercher à la vérifier expérimentalement, les points de désaccord entre les psychologues concernant surtout la localisation du filtre dans le modèle; Broadbent (1958) propose en effet une sélection précoce qui fait qu'elle précéderait l'analyse perceptuelle, alors que Deutsch & Deutsch (1963) ainsi que Norman (1968) soutiennent que celle-ci intervient plus tardivement, juste avant le choix d'une réponse.

[10] 12 phénomènes phonétiques x 2 exemples x 2 types de stimuli x 2 variétés linguistiques = 96 stimuli.

à produire toutes les combinaisons ou paires possibles: 12 phénomènes phonétiques x 6 combinaisons possibles de type de stimuli (Mot_1-Mot_1; Mot_1-Mot_2; Mot-Logatome; Logatome–Mot; $Logatome_1$–$Logatome_1$; $Logatome_1$–$Logatome_2$) x 4 combinaisons possibles de variétés linguistiques (VQ-VE; VE-VQ; VQ-VQ; VE-VE) pour un total de 288 combinaisons possibles distribuées en 6 jeux.

Tableau 1
Stimuli de l'expérience III

Phénomènes phonétiques	mot	logatome	mot	logatome
affrication de /d/	duc	dube	dîme	dife
affrication de /t/	tige	tive	tube	tupe
relâchement de /u/	poule	doule	soupe	foupe
relâchement de /y/	bûche	puche	jupe	chupe
relâchement de /i/	pipe	quipe	ville	chille
diphtongaison de /ɑ̃/	lampe	lanque	banque	bampe
diphtongaison de /ɛ̃/	singe	sinse	linge	lince
diphtongaison de /ɛ/	terre	tège	neige	nève
diphtongaison de /ɔ/	guêpe	quêpe	caisse	daisse
diphtongaison de /œ/	beurre	deure	coeur	teure
diphtongaison de /ɑ/	gaz	dase	base	pase
postériorisation de /ɑ/	drap	tra	plat	kla

4.2.2 Sujets

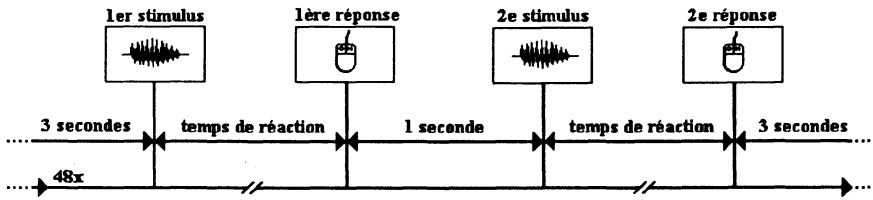
L'expérience a été menée auprès de 40 étudiants inscrits à un programme de baccalauréat en langue française ou en linguistique à l'Université Laval à Québec qui ignoraient l'objet de l'expérimentation. Tous les sujets avaient le français québécois comme langue maternelle, mais possédaient une connaissance au moins passive de la variété européenne, étant donné l'utilisation qui en est faite dans les médias, en particulier au cinéma, et éventuellement dans l'enseignement universitaire. Afin de contrôler les effets possibles de la spécialisation hémisphérique du cerveau, seuls des droitiers ont été retenus. L'âge moyen des sujets était de 23 ans, avec une étendue de 20 à 34 ans; la majorité d'entre eux étaient de sexe féminin (28 femmes et 12 hommes).

4.2.3 Protocole expérimental

Le test a été administré individuellement en salle anéchoïque. Comme pour l'expérience II, le déroulement du test était entièrement contrôlé par un micro-ordinateur compatible IBM de type AT équipé d'un convertisseur permettant de reproduire un signal numérique de façon analogique. Le sujet s'assoyait à une table sur laquelle se trouvait un écran cathodique et prenait en main une souris modifiée dont les deux boutons avaient été colorés. Pour la tâche de décision lexicale, on demandait au sujet d'appuyer sur le bouton de gauche, coloré en vert, lorsqu'il croyait percevoir un mot et d'appuyer sur le bouton de droite, en rouge, lorsqu'il jugeait que le stimulus perçu était un non-mot, c'est-à-dire un logatome. L'écran était utilisé uniquement pour informer le sujet sur le nombre de stimuli qu'il lui restait à identifier et pour lui rappeler l'utilisation des deux boutons de la souris. Pour ne pas distraire les sujets de la tâche qui leur était assignée, aucune autre information n'apparaissait à l'écran. À gauche et à droite, sur la table, deux enceintes acoustiques étaient placées à 75 cm du sujet et dirigées vers lui. Tous les autres appareils avaient été placés à l'extérieur de la salle afin de réduire le bruit et les possibilités de distraction.

Avant de commencer le test, le sujet s'identifiait, lisait la consigne affichée à l'écran, prenait en main la souris en plaçant les pouces sur les boutons et se familiarisait avec la procédure et l'équipement. La figure 3 illustre le déroulement séquentiel de l'expérimentation.

Figure 3
Protocole de l'expérience III



Chaque paire de stimuli était précédée d'une pause de trois secondes. L'ordinateur faisait entendre le premier stimulus sonore de la paire et attendait la pression de l'un des deux boutons de la souris par le sujet. Le temps de réaction à ce premier stimulus mesuré en millisecondes de même que la réponse étaient automatiquement enregistrés dans une base de données. Le second stimulus suivait une seconde plus tard et les deux variables (temps de réaction et réponse) étaient à leur tour inscrites dans la base de données. Cette procédure était répétée pour chacune des 48 paires de stimuli; la durée moyenne du test était d'un peu moins de six minutes.

Les données ainsi recueillies pour chaque sujet occupent 1920 entrées dans la base de données (48 paires x 40 sujets), lesquelles totalisent 3840 temps de réponses et autant de décisions lexicales. Une réponse a été considérée *anticipative* et marquée comme telle pour un traitement spécial quand son temps de réaction était inférieur à 90 millisecondes[11]. De la même façon, une réponse a été marquée comme constituant un *retard* quand elle était produite après un délai de plus 1500 ms pour un mot et de plus de 2000 ms pour un logatome. Ces seuils pour les retards ont été établis empiriquement une fois l'expérimentation complétée et correspondent à une distance de trois écarts types par rapport à la moyenne.

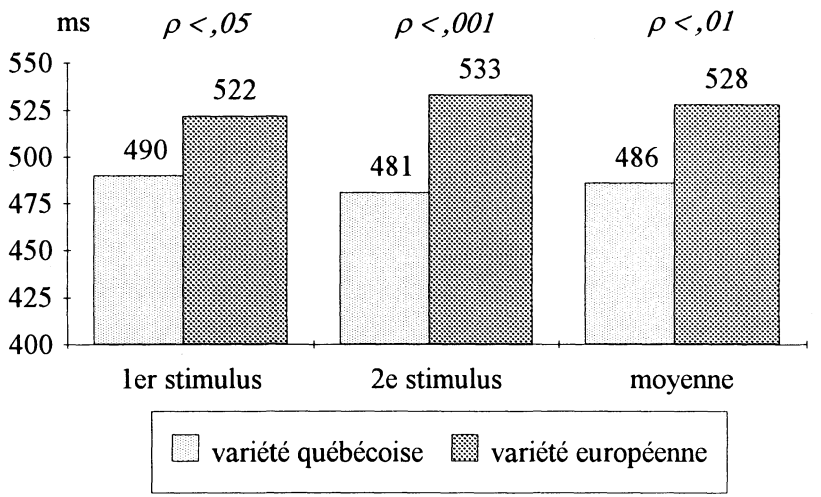
4.3 Résultats

4.3.1 Temps de réaction en fonction de la variété linguistique

Les moyennes des temps de réaction en fonction de la variété linguistique et de la position du stimulus dans la paire sont présentées à la figure 4.

[11] Il s'agit là du temps de réponse irréductible pour un court stimulus sonore tel qu'établi par Snodgrass, Levy-Berger & Haydon (1985, p. 101).

Figure 4
Moyennes des temps de réaction en fonction de la variété linguistique et de la position dans la paire



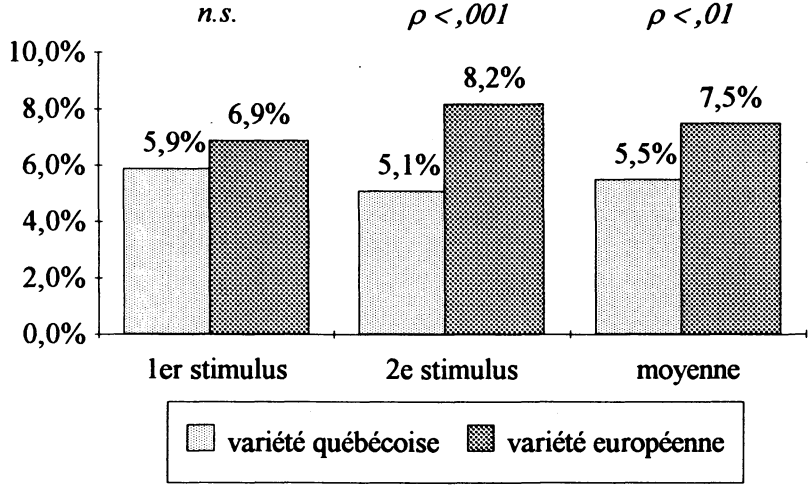
On y observe que le temps de réponse moyen aux stimuli en variété québécoise est dans l'ensemble, toutes positions confondues, sensiblement plus court que pour les stimuli en variété européenne (485 ms contre 527 ms), la différence étant significative à ,001. Cet avantage global de la variété vernaculaire se retrouve peu importe la position du stimulus dans la paire, même si l'on peut noter que la position a un effet différent selon la variété en cause, puisque l'on constate une diminution du temps de réaction pour les deuxièmes stimuli en variété québécoise, mais une augmentation pour ceux en variété européenne; ce phénomène met en cause l'effet possible du filtre sur lequel nous reviendrons.

4.3.2 Taux de décisions correctes en fonction de la variété linguistique

Ces premiers résultats semblent appuyer notre hypothèse sur la plus grande facilité de traitement de la variété familière. Ils doivent cependant être mis en parallèle avec les taux de réussite à la tâche de décision lexicale afin d'écarter l'effet possible de ce que les psycholinguistes appellent la *fonction de compensation vitesse/précision*, cf. Snodgrass & al. (1985, p. 103). Il convient en effet de s'assurer que les temps de réaction plus courts ne s'accompagnent pas d'un taux

d'erreur plus élevé dans la tâche, auquel cas on peut penser qu'il s'agit d'un simple déplacement de priorité en faveur de la vitesse au détriment de la précision. Si l'on considère les taux d'erreurs de décision[12] en fonction de la variété linguistique et de la position du stimulus qui sont présentés à la figure 5, on observe que les stimuli en variété québécoise ont donné lieu à moins d'erreurs de décision; cette corrélation positive entre la vitesse et le taux de réussite, en même temps qu'elle permet d'exclure une quelconque fonction de compensation[13], constitue un argument de poids en faveur de l'hypothèse d'une plus grande facilité et d'une plus grande efficacité de traitement lorsque l'on a affaire à la variété linguistique plus familière.

Figure 5
Pourcentages d'erreurs de décision lexicale en fonction de la variété linguistique et de la position du stimulus



[12] Les erreurs de décision comprennent toutes les erreurs d'identification du statut d'un stimulus (mot/non-mot), mais n'incluent pas les réponses anticipatives et les retards.

[13] Cette nécessité d'une corrélation entre le temps de réaction et le taux de réussite à la tâche est clairement posée dans Snodgrass & al. (1985, p. 104): «As long as error rates are positively correlated with RT [Reaction Time], so that low error rates accompany fast RTs and high error rates accompany slow RTs, the investigator may feel fairly confident that the pattern of RTs cannot be accounted for by speed-accuracy trade-off. On the other hand, if error rates are negatively correlated with RT, then a speed-accuracy trade-off is a possible alternative explanation to the investigator preferred hypothesis.»

Le fait que les réponses anticipatives soient plus nombreuses du côté des stimuli en variété québécoise (2,7% contre 1,6%) alors que les retards se retrouvent principalement en variété européenne (1,2% contre 0,8%) représente encore un indice supplémentaire de la plus grande aisance des sujets lorsque les stimuli étaient présentés dans la variété vernaculaire.

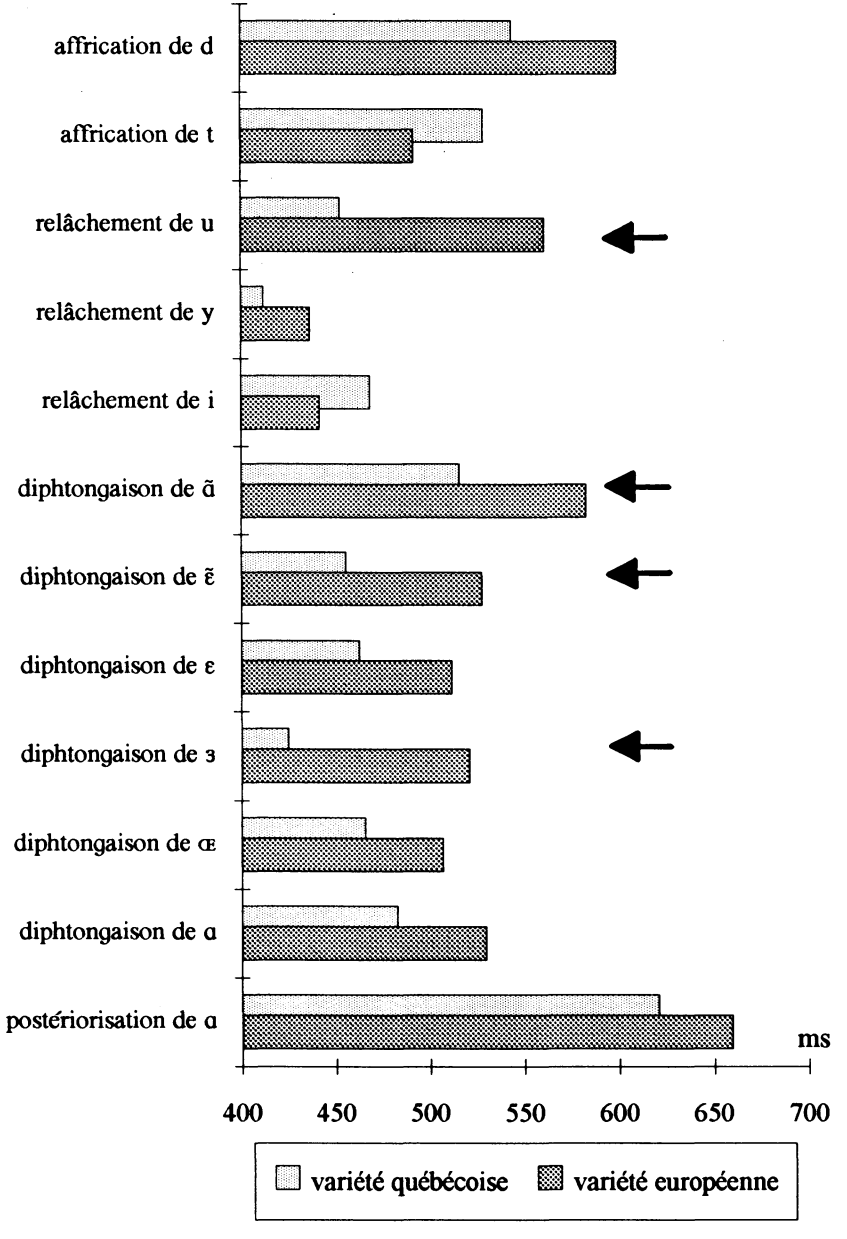
4.3.3 Effets de la variété linguistique en fonction des divers phénomènes phonétiques

Il peut être intéressant, aussi bien dans une perspective sociophonétique que dans une optique d'optimisation des systèmes de synthèse à un coût minimal, de se demander dans quelle mesure cet avantage global de la variété québécoise se retrouve pour chacun des douze phénomènes phonétiques pris en compte dans cette expérience. La figure 6 montre à cet effet que le temps de réaction moyen est moins élevé pour la forme québécoise dans 10 cas sur 12 (les deux phénomènes qui font exception étant l'affrication de /t/ et le relâchement de /i/; les différences significatives à ,05 sont signalées par une flèche).

La même tendance se laisse observer lorsque l'on prend en considération les taux d'erreurs de décision, ceux-ci étant plus faibles avec la variété québécoise dans tous les cas, sauf pour ce qui est de la diphtongaison du /ɑ/ devant /z/ ou /r/ et, encore une fois, de l'affrication de /t/. On peut donc conclure que, pour la très grande majorité des phénomènes phonétiques examinés, l'utilisation de la variété familière a un effet facilitant sur le traitement. L'existence de résultats divergents pour des phénomènes étroitement apparentés (l'affrication du /t/ par rapport à celle du /d/ ou le comportement différent du /i/ par rapport aux autres voyelles hautes) soulève cependant un problème auquel il est difficile d'apporter une réponse en termes phonétiques ou sociophonétiques. Peut-être faut-il y voir une simple variation statistiquement possible à l'intérieur de l'intervalle de confiance des moyennes[14], sans écarter non plus la possibilité d'un manque d'homogénéité dans la réalisation phonétique des stimuli; il est en effet extrêmement difficile – et c'est là une lacune du protocole expérimental dont nous sommes conscients – de s'assurer que les traits phonétiques que l'on souhaite tester sont réalisés de façon similaire dans tous les

[14] Cet intervalle de confiance à 95% des moyennes est de 440-545 ms en VE et de 479-578 ms en VQ dans le cas de l'affrication du /t/ alors qu'il est 398-486 ms (VE) et 434-504 ms (VQ) pour le relâchement du /i/; il y a donc possibilité de recouvrement des moyennes de VE et de VQ dans les deux cas.

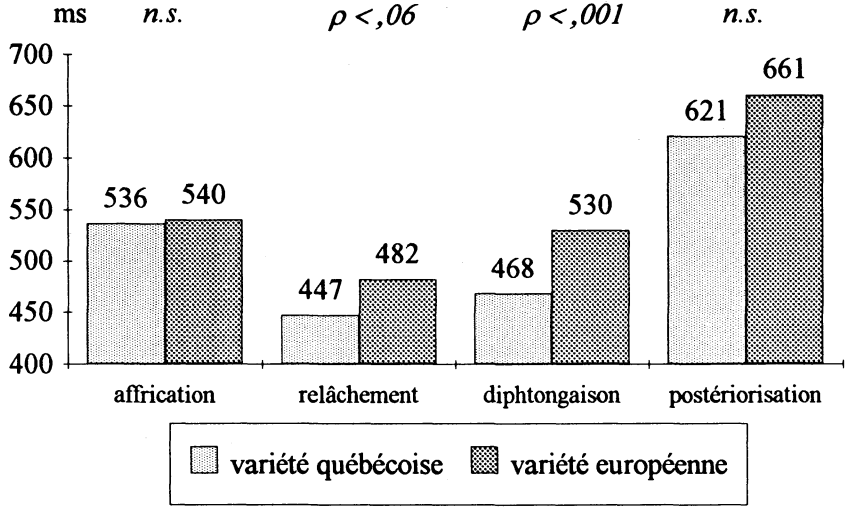
Figure 6
Moyenne des temps de réaction pour chacun des phénomènes phonétiques en fonction de la variété linguistique utilisée



stimuli, une réalisation trop accentuée ou au contraire trop discrète pouvant avoir un effet perturbateur important.

Le fait d'opérer un regroupement des douze phénomènes en quatre grandes classes, soit l'affrication, le relâchement des voyelles hautes, la diphtongaison et finalement la postériorisation du /ɑ/ final, minimise ces inconvénients et facilite le traitement statistique. Les résultats pour un tel regroupement sont présentés à la figure 7.

Figure 7
Moyennes des temps de réaction pour les types de phénomènes phonétiques en fonction de la variété linguistique utilisée



On y constate encore un fois, dans trois cas sur quatre, un avantage sensible en faveur de la variété québécoise, mais ce n'est que pour la diphtongaison que la différence se révèle statistiquement significative. Si l'absence de différence significative pour l'affrication et le relâchement n'est pas sans relation avec le caractère divergent des résultats pour le /t/ et le /i/ observé précédemment, il n'est pas sans intérêt par ailleurs de rappeler les conclusions de Lappin (1982) et de Tremblay (1990) à l'effet que l'affrication de /t/ et de /d/ de même que le relâchement des voyelles hautes sont parmi les traits typiques du français québécois qui passent le

plus inaperçus alors que les phénomènes de diphtongaison sont encore stigmatisés[15].

4.3.4 Effets du filtre linguistique sur les temps de réaction

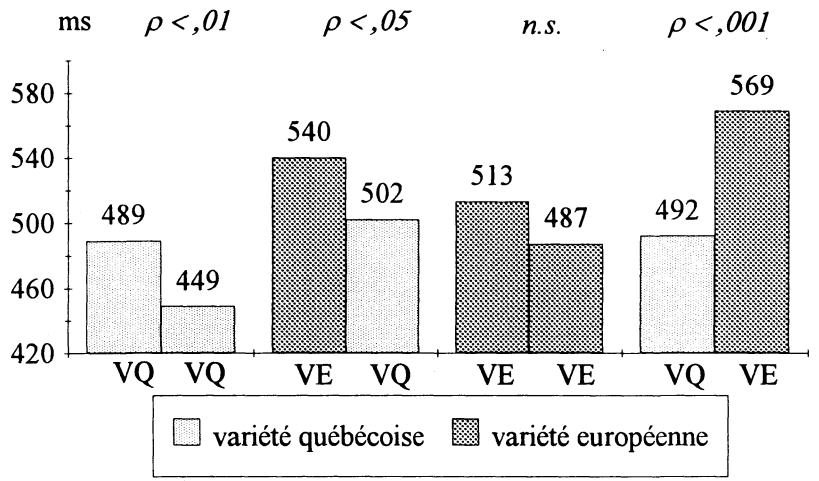
L'expérience III repose pour une large part sur l'hypothèse que la variété linguistique utilisée fonctionne comme un filtre susceptible de favoriser ou de gêner le traitement du stimulus. Le comportement différent observé pour le deuxième stimulus (cf. ci-dessus 4.3.1) selon qu'il était produit en variété québécoise (diminution du temps de réponse) ou en variété européenne (augmentation du temps de réponse) suggérait déjà, comme on l'a signalé, un effet de filtre. Cet effet de filtre ressort nettement si on examine le détail de la figure 8 qui présente les temps moyens de réponse en fonction de la position du stimulus, de la variété linguistique du stimulus (VQ ou VE) et de la variété linguistique du stimulus précédent (VQ ou VE). Les diverses combinaisons possibles sont dans l'ordre VQ-VQ (variété québécoise sans changement de filtre), VE-VQ (changement de filtre de la variété européenne vers la variété québécoise), VE-VE (variété européenne sans changement de filtre) et finalement VQ-VE (changement de filtre de la variété québécoise vers la variété européenne).

Ces résultats confirment la tendance observée jusqu'à maintenant, à savoir que les stimuli en variété québécoise sont toujours, à conditions comparables (position dans la paire et rapport au stimulus précédent), plus rapidement traités que ceux en variété européenne. On remarque aussi que la nature du premier stimulus semble avoir un effet notable, et statistiquement significatif dans trois cas sur quatre, sur la facilité de traitement du second. C'est en effet dans le cas où le second stimulus est précédé d'un stimulus dans la même variété que l'on enregistre les meilleurs résultats, et cela aussi bien pour la variété québécoise (449 ms) que pour la variété européenne (487 ms). Inversement, c'est lorsqu'il succède à un premier stimulus dans l'autre variété, que les temps de réponse pour le second sont les plus élevés en variété québécoise (502 ms) comme en variété européenne (569 ms). De tels résultats sont cohérents avec l'hypothèse voulant que la variété linguistique agisse comme un filtre sur la perception et le traitement de stimuli linguistiques. Les temps de réponse plus longs relevés lorsque le second stimulus n'est pas dans la même variété que le premier peuvent alors s'expliquer par le temps requis pour la

[15] La situation est plus confuse pour la postériorisation du /ɑ/, une postériorisation modérée semblant passer inaperçue alors que la postériorisation extrême fait l'objet d'une évaluation négative.

mise en place du filtre approprié, travail supplémentaire dont il est évidemment fait économie lorsque les deux stimuli sont dans la même variété.

Figure 8
Moyennes des temps de réaction en fonction de la variété linguistique et du filtre mis en place



Un examen un peu plus attentif des faits conduit cependant à distinguer deux dimensions distinctes de cet effet de filtre. Si l'amélioration des performances pour le second stimulus lorsque les deux stimuli de la paire appartiennent à la même variété suggère que la différence peut être imputée à l'économie du temps requis pour la mise en place du filtre, l'avantage général de la variété québécoise, à conditions comparables, et même dans le contexte dévaforable VE-VQ, amène à penser que la variété québécoise constitue, pour des locuteurs québécois, un filtre intrinsèquement plus efficace et plus performant que la variété européenne, ce qui n'a rien d'étonnant si on se rappelle les enseignements de la sociolinguistique sur le caractère plus régulier et plus systématique du vernaculaire, cf. Labov (1976, p. 289). Il semble en outre que les effets de *filtre naturel* associés à la variété linguistique la plus familière et ceux de *filtre contextuel* découlant de la concordance ou non avec le filtre mis en place pour le traitement du stimulus qui précède soient additifs, puisque c'est dans le contexte VQ-VQ que l'on obtient le temps de réaction

moyen le plus court alors que le temps le plus long est observé dans le contexte VQ-VE.

L'amélioration notable des performances pour le second stimulus de la paire VQ-VQ (-40 ms par rapport au premier) soulève cependant certaines questions dans une perspective sociolinguistique dans la mesure où elle pourrait suggérer que la variété vernaculaire n'est pas aussi *naturelle* qu'on pourrait le croire puisqu'il résulte un gain important de la mise en place préalable du filtre approprié (premier stimulus en VQ). Plutôt que de penser que le locuteur dispose en permanence d'un filtre adapté à la variété linguistique de la communauté dans laquelle il s'inscrit, faudrait-il alors se tourner vers un modèle de *tabula rasa* qui suppose qu'un filtre ou l'autre doit de toute façon être mise en place? La réponse à cette question n'est pas simple; divers éléments d'explication peuvent être avancés, les uns de nature psycholinguistique (le filtre, même *naturel*, demandant à être activé), les autres de caractère sociolinguistique (le «bilinguisme» relatif des sujets québécois qui possèdent au moins passivement la variété européenne); d'autres enfin mettent en cause les caractéristiques même du protocole expérimental qui, en plus de constituer une situation formelle où le vernaculaire n'est pas attendu, repose expressément sur l'alternance de stimuli de l'une et l'autre variété. Nous ne sommes pas en mesure de faire le départage des rôles respectifs de ces divers facteurs, mais il est certain que l'effet des conditions d'expérimentation doit être pris en considération avant de tirer quelque conclusion que ce soit.

5. Conclusion

Tous ces résultats qui font voir, tout spécialement dans le cadre de l'expérience III, un avantage très net pour la variété familière en termes de vitesse de traitement comme de précision de la réponse rejoignent évidemment des positions assumées depuis longtemps en sociolinguistique quant au statut du vernaculaire dans le rapport du sujet à sa langue. Ils n'en présentent pas moins, croyons-nous, un intérêt nouveau dans la mesure où ils constituent une avancée dans le champ encore peu exploré des problèmes de perception au terme d'une démarche qui a emprunté beaucoup aux études de psychologie et de psycholinguistique. Ils témoignent ce faisant de la possibilité pour la sociolinguistique de trouver de nouveaux champs d'application où elle sera mise à contribution, en conjonction avec d'autres disciplines, pour la solution de problèmes pratiques comme celui qui a retenu notre attention.

Cette mise en lumière d'une plus grande facilité de traitement de la variété familière ne suffit évidemment pas à elle seule à justifier le développement de systèmes de synthèse pouvant intégrer diverses variétés, puisqu'il reste à s'assurer que cet avantage en situation expérimentale impliquant des mots isolés se trouverait conservé dans des conditions d'utilisation concrète faisant appel à la parole continue. Dans la mesure où ils mettent en cause la dimension fondamentale de l'intelligibilité, nos résultats appellent résolument à la poursuite des recherches dans cette direction et font voir que les enjeux de la prise en compte des variétés linguistiques ne sont pas uniquement idéologiques, la préservation de l'identité culturelle et la légitimation de la variété vernaculaire pouvant aller de pair avec une amélioration de la perfomance des systèmes.

Claude Paradis
C.I.R.A.L., Université Laval

Martin Brousseau
Université Laval
et Université de Provence

Jean Dolbec
Université du Québec à Chicoutimi

Références

BLADON, A., R. CARLSON, B. GRANSTRÖM, S. HUNNICUTT & I. KARLSSON (1987) «A text-to-speech system for British English, and issues of dialect and style», in J. Laver & M. A. Jack (réd.), *European Conference on Speech Technology*, Edinburgh, CEP Consultants Ltd, vol. 1, pp. 55-58.

BROADBENT, D. E. (1958) *Perception and Communication*, Oxford, Oxford University Press.

BROUSSEAU, M. (1992) *Effets de l'utilisation de la variété québécoise sur l'intelligibilité de la parole de synthèse*, mémoire de maîtrise inédit, Université Laval, Québec.

CARLSON, R. (1992) «Synthesis: Modeling variability and constraints», *Speech Communication*, vol. 11, nos 2-3, pp. 159-166.

DEUTSCH, J. A. & D. DEUTSCH (1963) «Some theorical considerations», *Psychological Review*, no 70, pp. 80-90.

DOLBEC, J. & Cl. PARADIS (1990) «Pour une prise en considération de la variation sociolectale dans la parole de synthèse», *Actes des XVIIIes Journées d'études sur la parole*, Montréal, Université de Montréal et Société française d'acoustique, pp. 40-44.

DOLBEC, J. & Cl. PARADIS (1991) «Modélisation de la variation phonétique en français québécois», *Revue québécoise de linguistique théorique et appliquée*, vol. 10, no 3, pp. 85-105.

DOLBEC, J., Cl. PARADIS, M. BROUSSEAU & D. DESHAIES (1991) «Évaluation de l'effet de la variété linguistique utilisée sur l'intelligibilité de la parole synthétique en contexte québécois», *Actes du XIIe Congrès international des sciences phonétiques*, Aix-en-Provence, Publications de l'Université de Provence, vol. 3, pp. 486-489.

FANT, G. (1991) «What can basic research contribute to speech synthesis?», *Journal of Phonetics*, vol. 19, no 1, pp. 75-90.

KARLSSON, I. (1991) «Female voice in speech synthesis», *Journal of Phonetics*, vol. 19, no 1, pp. 111-120.

LABOV, W. (1976) *Sociolinguistique*, Paris, Éditions de Minuit.

LAPPIN, K. (1982) «Évaluation de la prononciation du français montréalais: étude socio-linguistique», *Revue québécoise de linguistique*, vol. 11, no 2, pp. 91-112.

LEE, K.-F. (1989) «Hidden Markov models: Past, present and future», in J. P. Tubach & J. J. Mariani (réd.), *Eurospeech 89. European Conference on Speech Communication and Technology*, Edinburgh, CEP Consultants Ltd, vol. 1, pp. 145-155.

MÉAR-CRINE, A. & T. LECLERC (1976) «Attitudes des adolescents canadiens-français vis-à-vis du franco-québécois et du français académique», *Cahiers de linguistique*, n° 6, pp. 155-170.

MEYER, D. E., R. W. SCHVANEVELDT & M. G. RUDDY (1975) «Loci of contextual effects on visual word recognition», in P. M. A. Rabbitt & S. Dornic (réd.), *Attention and Performance* V, London, Academic Press, pp. 98-118.

NORMAN, D. A. (1968) «Towards a theory of memory and attention», *Psychological Review*, vol. 75, pp. 522-536.

PARADIS, C. & J. DOLBEC (1992) *PHONO. Applicateur de règles phonologiques*, Québec, C.I.R.A.L., Université Laval.

PAVLOVIC, C. V., M. ROSSI & R. ESPESSER (1990) «Use of the magnitude estimation technique for assessing the performance of text-to-speech synthesis systems», *Journal of the Acoustical Society of America*, vol. 87, n° 1, pp. 373-382.

ROBERT, J.-M., A. CHOINIÈRE & R. DESCOUT (1989) «Subjective evaluation of the naturalness and acceptability of three text-to-speech systems in French», in J. P. Tubach & J. J. Mariani (réd.), *Eurospeech 89. European Conference on Speech Communication and Technology*, Edinburgh, CEP Consultants Ltd, vol. 2, pp. 640-643.

SNODGRASS, J. G., G. LEVY-BERGER & M. HAYDON (1985) *Human Experimental Psychology*, New York, Oxford University Press.

TREMBLAY, L. (1990) «Attitudes linguistiques et perception sociale de variables phonétiques», *Revue québécoise de linguistique théorique et appliquée*, vol. 9, n° 3, pp. 197-221.

TRUDGILL, P. (1974) *Sociolinguistics*, Harmondsworth, Penguin Books Ltd.