

Les systèmes vocaliques des dialectes de l'anglais britannique

Emmanuel Ferragne, François Pellegrino

Laboratoire Dynamique Du Langage
UMR CNRS 5596 – Université Lyon 2
14 avenue Berthelot
69007 LYON
Emmanuel.Ferragne@univ-lyon2.fr
<http://www.ddl.ish-lyon.cnrs.fr>

ABSTRACT

This paper is an attempt to characterize the vowel systems of the dialects of British English. We carried out a (semi-) automatic dialect identification procedure using [hVd] words. Our second aim was to examine to what extent the procedure allowed the description of vowel systems. The method yields approximately 90% correct identification, and we show that it is not sensitive to gender differences, and may therefore be used for the description of vowel systems.

1. INTRODUCTION

Les dialectes de l'anglais britannique ont été abondamment décrits dans la littérature traditionnelle d'avant l'ère de la phonétique de corpus (Wells [1]). Les ouvrages de référence récents, quoique basés sur l'analyse minutieuse de corpora recueillis *ad hoc* (Kortmann & Schneider [2], Foulkes & Docherty [3]), se concentrent davantage sur la variation sociolinguistique à l'intérieur d'un même dialecte, si bien que la mise en lumière des caractéristiques de prononciation régionales actuelles – au-delà de certains schibboleths bien connus – n'est pas aisée. Il semblerait même que les études contemporaines aient tendance à considérer la variation sociolinguistique seule comme digne d'intérêt, reléguant la variation régionale au statut de curiosité folklorique (voir par exemple le projet *Voices* de la BBC¹).

Les bases de données d'enregistrements des dialectes britanniques récentes et disponibles pour la recherche sont, à notre connaissance, au nombre de deux : le corpus IViE², conçu pour l'étude de l'intonation, et le corpus Accents of the British Isles (ABI – D'Arcy *et al.* [4]), que nous utiliserons dans cette étude.

Notre analyse consiste à procéder à une identification (semi-) automatique du dialecte basée sur l'information contenue dans les systèmes vocaliques des variétés régionales de notre corpus. Parallèlement à l'identification automatique, nous aborderons des aspects plus phonologiques, notamment, les convergences et scissions phonémiques.

Les différences systémiques les plus diagnostiques en Angleterre proviennent de la scission phonémique de FOOT-STRUT et du phénomène appelé BATH-Broadening. Le premier fait référence au fait qu'avant le 17^e siècle, *luck* et *look* étaient de parfaits homophones (avec un timbre proche de [u]). A cette époque, une délabialisation de la voyelle de *luck*, amorcée dans le sud de l'Angleterre a conduit à une phonémisation de l'opposition FOOT-STRUT au sud (Wells [1]). Ce phénomène ne s'est pas diffusé au nord, si bien que l'absence de scission y constitue un trait caractéristique, souvent stigmatisé. Le BATH-Broadening est également une innovation du sud de l'Angleterre ; la voyelle s'est allongée devant fricative au 17^e siècle et a acquis une qualité postérieure au 19^e. Certes, le statut phonémique de l'opposition BATH-TRAP demeure discutable – au reste, cette opposition n'a pas pu être étudiée dans notre corpus – mais elle méritait d'être mentionnée car elle constitue, au même titre que FOOT-STRUT une frontière nord/sud bien connue du dialectologue et du locuteur moyen. Wells [1] et Beal [5] notent l'existence d'une convergence NURSE-SQUARE à Liverpool ; c'est également le cas à Hull (Williams & Kerswill [6]), ville où ont eu lieu les enregistrements du dialecte étiqueté "East Yorkshire" dans notre corpus. Une description sommaire de l'anglais écossais fait apparaître une absence de contraste possible entre les voyelles de GOOSE et FOOT et celles de LOT et THOUGHT. Un examen plus détaillé et récent (Stuart-Smith [7]) tend à confirmer ces convergences phonémiques (au moins à Glasgow) tout en soulignant qu'elles varient en fonction de facteurs sociolinguistiques. FOOT et STRUT sont, quant à eux bien distincts, et le BATH-Broadening n'a pas eu lieu. L'anglais du Pays de Galles, tel qu'il est décrit chez Wells [1] et Penhallurick [8] ne fait pas apparaître de convergence ou scission phonémique que nous puissions tester. En effet, Wells [1] mentionne la convergence STRUT-Schwa (ici encore, le statut phonémique est contestable), ce que nous ne pouvons pas étudier puisque nous ne disposons que de voyelles accentuées et que les membres de l'ensemble lexical Schwa apparaissent en position inaccentuée. Penhallurick [8] mentionne toutefois la possibilité d'une réalisation [u] pour STRUT dans le nord-est du pays, ce qui nous laisserait peut-être envisager l'existence d'une convergence FOOT-STRUT.

¹ <http://www.bbc.co.uk/voices/>

² <http://www.phon.ox.ac.uk/~esther/ivyweb/>

2. CORPUS ET MÉTHODE

2.1. Description du corpus

Nous avons utilisé les mots en [hVd] du corpus Accents of the British Isles (ABI). Le corpus ABI est une base d'enregistrements de parole lue effectués en 2003 et qui regroupe 14 dialectes des Iles Britanniques, pour un total de 284 locuteurs. En moyenne, chaque dialecte est représenté par dix locuteurs masculins et dix locutrices féminines. Chaque locuteur répète cinq fois une liste de 19 mots, dont certains sont des non-mots, à structure [hVd]. Chacune des 19 voyelles employées constitue une opposition potentielle dans le système phonologique. En d'autres termes, les dialectes possédant l'inventaire vocalique le plus riche ont jusqu'à 19 phonèmes. Les fichiers son sont au format Windows PCM, 22050 Hz, 16 bits, mono. Les enregistrements ont eu lieu dans des salles calmes (ex : salles de bibliothèques publiques) avec un micro-casque directement relié à un PC via une carte son externe. La liste des dialectes et les abréviations correspondantes est donnée dans la table 1. Ces enregistrements sont majoritairement de qualité moyenne en cela qu'ils comportent certains bruits environnants, des interventions de la personne qui effectue l'enregistrement, et que la compétence des locuteurs de certains dialectes en lecture est – aux dires des phonéticiens anglophones qui ont eu l'occasion de les entendre – très décevante. De plus, si l'on se place dans une optique de dialectologie moderne (dominée par la sociolinguistique), le corpus est inexploitable puisque aucune donnée n'est fournie concernant l'âge des locuteurs, leur occupation et leur histoire linguistique.

Table 1 : liste des dialectes.

abréviation	étiquette du dialecte
brm	Birmingham
crn	Cornouailles
ean	East-Anglia
eyk	East-Yorkshire
gla	Glasgow
lan	Lancashire
lvp	Liverpool
ncl	Newcaslité
nwa	North Wales
roi	Republic of Ireland
shl	Scottish Highlands
sse	Standard Southern English
uls	Ulster

2.2. Pré-traitement des données

Le corpus ABI est segmenté au niveau du mot. A partir des frontières délimitant chaque mot en [hVd], nous avons procédé à une détection automatique du voisement (avec la librairie Snack en Tcl/Tk³) ; nous avons alors considéré que ces trames voisées constituaient la partie vocalique du mot. A partir des

voyelles ainsi isolées, 12 MFCC plus l'énergie ont été calculés à l'aide de la toolbox Rasta⁴ pour Matlab avec un pas d'analyse de 10 ms, une fenêtre de 20 ms. Chaque voyelle est ensuite représentée par un vecteur de 52 valeurs : 12 MFCC et l'énergie à $\frac{1}{4}$, $\frac{1}{2}$ et $\frac{3}{4}$ de la durée, les deltas MFCC et la dérivée de l'énergie calculés sur 5 trames centrées sur la trame du milieu temporel ainsi que, optionnellement, une 53^e valeur : la durée de la voyelle.

2.3. Classification

La méthode de classification utilisée ici s'inspire largement des travaux de Huckvale [9]. Elle consiste à obtenir une représentation du système vocalique de chaque locuteur à partir d'une matrice de distances euclidiennes entre les voyelles de ce locuteur. Ceci constitue en soi une forme de normalisation du locuteur puisque l'on ne compare pas directement les voyelles d'un locuteur avec celles d'un autre, mais les voyelles d'un même locuteur entre elles. Pour un même locuteur et un timbre donné, ce sont les valeurs moyennes qui ont été employées pour l'analyse et non pas les valeurs individuelles de chaque répétition. On calcule donc la matrice de distances entre N voyelles pour chaque locuteur, puis la matrice de distance moyenne pour chaque dialecte. La méthode de validation est basée sur le principe du "leave-one-out" : la matrice de distances entre segments de chaque locuteur est comparée à la matrice (moyenne) de chaque dialecte – dont celle de son dialecte calculée sans la matrice du locuteur en cours de test. Ce dernier est classé comme appartenant au dialecte dont la matrice se rapproche le plus de la sienne. La similarité entre matrices est estimée grâce au coefficient de corrélation de Mantel. L'utilisation d'une corrélation entre matrices rend cette mesure insensible aux distances absolues entre les voyelles d'un locuteur. En outre, l'un des avantages de cette méthode de classification réside dans le fait qu'elle nécessite un temps de calcul extrêmement bref (moins de 7 secondes en temps CPU pour 145 locuteurs hommes sur un PC équipé de Windows XP et d'un processeur Intel Pentium 4, 3.2 GHz).

2.4. Pouvoir discriminant des différents timbres

Nous avons également souhaité tester le potentiel d'un nombre plus restreint de voyelles afin de faire apparaître les timbres les plus discriminants. Pour ce faire, la méthode de classification a été appliquée à chaque combinaison de 3 voyelles parmi 19, puis, la combinaison donnant le taux d'identification du dialecte le plus élevé a été conservée, et enrichie d'une des 16 voyelles restantes ; après avoir testé ces 16 ensembles de 4 voyelles, on garde le meilleur et on forme à partir de ce dernier 15 nouveaux ensembles, et ainsi de suite jusqu'à qu'il ne subsiste qu'un ensemble de 19 voyelles.

³ <http://www.speech.kth.se/snack/>

⁴ <http://www.ee.columbia.edu/labrosa/matlab/rastamat/>

Notons enfin que nous avons fait juger la typicalité de chacun des 145 locuteurs hommes du corpus sur la base d'un passage lu d'environ 90 mots par un phonéticien britannique expert. Cette analyse auditive préalable nous a conduit à exclure le dialecte portant l'étiquette "Inner London" car il était constitué de locuteurs immigrés d'ascendances très variées formant un tout trop hétérogène pour être considéré comme une seule et même entité.

3. RÉSULTATS ET DISCUSSION

3.1. Le système vocalique à onze monophthongues

En première approximation, nous avons classé les dialectes (sans tenir compte de la durée des voyelles) à partir des onze monophthongues de l'anglais britannique standard. Les trois modèles – hommes, femmes, sexes confondus – donnent respectivement des taux d'identification du dialecte de 80,7 ; 80,5 et 84%. Nous savons que la durée vocalique joue un rôle très important dans le système de la plupart des dialectes de l'anglais britannique (sauf, très vraisemblablement en Ecosse, voir Wells [1]) ; nous avons donc ajouté la durée aux paramètres spectraux précédents et avons pondéré empiriquement la durée par rapport à ces derniers. L'inclusion de la durée dans les modèles permet d'atteindre, 83,7 ; 87,5 et 87,8% dans les conditions homme, femmes, et sexes confondus respectivement.

3.2. Prise en compte des diphtongues

Nous avons ensuite tenté de déterminer les meilleures combinaisons de $p = 3$ à 19 voyelles parmi 19 (voir 2.4 pour les méthodes ; la durée n'est pas incluse). Les taux d'identification en fonction du nombre de voyelles sont donnés dans la figure 1.

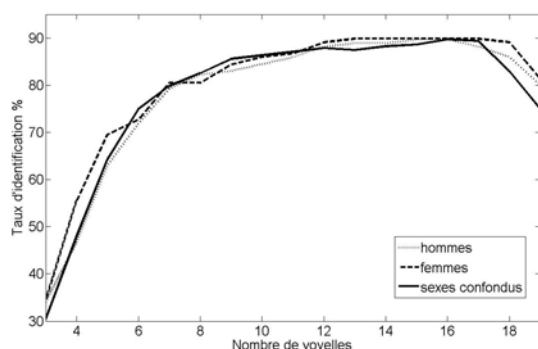


Figure 1 : taux d'identification du dialecte en fonction du nombre de voyelles incluses dans le modèle.

Le détail des meilleures combinaisons de voyelles apparaît dans la table 2. Les meilleurs taux pour les hommes, les femmes et sexes confondus sont de 89,6 ; 89,8 et 89,7% avec, respectivement, 15, 15 et 16 voyelles.

Table 2 : liste incrémentale des meilleures combinaisons de $p = 3$ à 19 voyelles parmi 19.

nb de voyelles	3	4	5	6	7	8	9	10	11	12	13	14	15	16
hommes	hid, hood, Hudd	heard	hide	whod	hade	hoed	hard	heed	hoard	hared	hod	head	had	.
femmes	hid, heard, hide	had	Hudd	hod	howd	head	whod	heered	hoild	hured	heed	hoard	hard	.
sexes confondus	hid, hood, Hudd	heard	hide	hade	whod	hoed	howd	hured	heered	heed	hod	head	hoard	had

Nous avons enfin inclus la durée et l'avons pondérée pour tenter d'améliorer les taux d'identification donnés dans la figure 1 à partir des combinaisons de voyelles de la table 2. Les taux d'identification ainsi obtenus sont de 89,6 ; 91,4 et 91,6% pour les conditions hommes, femmes et sexes confondus respectivement. La prise en compte de la durée ne permet d'améliorer le taux d'identification des femmes et des deux sexes confondus que dans une moindre mesure ; elle n'a en revanche aucun effet sur le modèle des hommes : cela est probablement dû à l'utilisation de mots en [hVd] hyper-articulés.

3.3. Discussion

Nous citerons, à titre de comparaison, deux études qui présentent des valeurs de formants pour l'identification du dialecte. Huckvale [9], qui a travaillé sur une autre partie du corpus que nous utilisons ici, partage ses voyelles en deux, utilise la valeur médiane des quatre premiers formants dans chaque moitié, et ses valeurs sont ensuite centrées-réduites par locuteur. Le vecteur moyen d'un timbre donné pour chaque locuteur et comparé aux vecteurs moyens de ce même timbre pour chaque dialecte. Dans la condition "tous sexes", il obtient 71,9% d'identification. Barry *et al.* [10] utilisent une méthode de classification basée sur des critères phonétiques et obtiennent environ 74% de classification correcte pour leurs 58 locuteurs en 4 aires dialectales grossières.

Dans la figure 1, nous constatons que la courbe d'identification tend à se stabiliser lorsque 7 voyelles sont prises en compte dans le modèle. Parmi ces 7 voyelles, la table 2 révèle que la paire *hood-Hudd*, qui correspond à la scission FOOT-STRUT mentionnée plus haut comme trait extrêmement diagnostique, est en bonne position pour les hommes et tous sexes confondus. La présence de *hade* n'est pas fortuite non plus puisque la diphtongaison (Ferragne & Pellegrino [11]) et un trait hautement discriminant. Le fait de retrouver cette voyelle aussi bien placée dans la table 2 tend à prouver que notre méthode capture de façon adéquate les caractéristiques dynamiques de voyelles.

Notons encore qu'un des traits phonétiques que nous avons mesuré n'est pas à proprement parler vocalique : il s'agit de la rhoticité, i.e. la réalisation ou non d'un /r/ postvocalique, et la nature articulatoire de ce /r/. Par exemple, le sud-est de l'Angleterre n'est pas rhotique, le sud-ouest a un /r/ souvent rétroflexe (Altendorf & Watt [12]) ; au Pays de Galles, le /r/ peut être une approximante post-alvéolaire, un trille ou une battue (Penhallurick [8]). Il est donc certain que, outre les différences des systèmes vocaliques, notre

classification a bénéficié de la variation géographique du /r/.

Comme l'a noté Huckvale [9], nos résultats confirment que cette méthode est insensible aux différences de sexe. Elle semble d'ailleurs tout a fait adéquate pour représenter des systèmes vocaliques en évitant d'avoir recours à une normalisation du locuteur fastidieuse et souvent inefficace basée sur des valeurs de formants. L'utilisation des MFCC permet une procédure entièrement automatisée alors que l'extraction automatique de formants nécessite souvent l'intervention d'un expert humain. Afin d'illustrer l'intérêt de notre méthode pour la description phonétique, la figure 2 représente un dendrogramme obtenu à partir de la matrice de distance moyenne des 11 monophthongues du dialecte East Yorkshire. On visualise aisément l'absence de scission FOOT-STRUT (*hood* et *Hudd* sur la figure), typique des dialectes du nord de l'Angleterre, et la proximité relative des voyelles de *heed* et *hid*, de *head* et *heard*, de *had* et *hard*, et de *hod* et *hoard*.

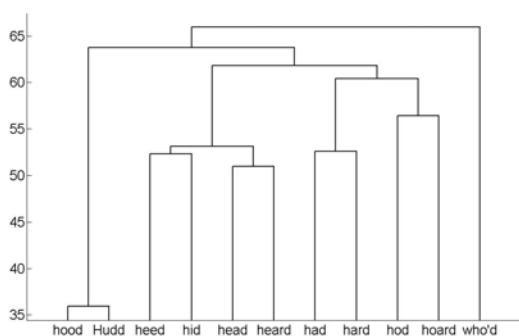


Figure 2 : Classification hiérarchique des voyelles du dialecte East Yorkshire.

En ce qui concerne les résultats de l'identification, nous pouvons les comparer directement avec ceux de Huckvale [9] : ce dernier avait obtenu 86,9% dans la condition tous sexes confondus, et 87,2% pour la condition "même sexe" en comparant des matrices de distances de 140 voyelles issues de phrases lues. Il est intéressant de noter que malgré le fait que les mots en [hVd] soient communément considérés comme trop éloignés de la parole spontanée, notre étude prouve qu'ils contiennent suffisamment d'information pour la description et l'identification de certains dialectes.

4. CONCLUSION

A partir des MFCC extraits des mots en [hVd] du corpus ABI, nous obtenons un taux d'identification du dialecte de 89,6 ; 91,4 et 91,6% pour les hommes, femmes et tous sexes confondus respectivement. Le taux d'identification à partir des MFCC des monophthongues (84%) est amélioré lorsque la durée est incluse dans le modèle (87,8%), et l'ajout des diphtongues permet d'atteindre 91,6%. La méthode des matrices de distances entre timbres est ainsi robuste par

rapport aux différences de sexe. Ceci permet une représentation fiable des systèmes vocaliques des différents dialectes de l'anglais britannique.

BIBLIOGRAPHIE

- [1] J.C. Wells. *Accents of English*. Cambridge University Press, Cambridge, UK, 1982.
- [2] B. Kortmann and E.W. Schneider. *A Handbook of Varieties of English*. Mouton de Gruyter, Berlin, 2004.
- [3] P. Foulkes and G. Docherty. *Urban Voices. Accent Studies in the British Isles*. Arnold, Londres, 1999.
- [4] S.M. D'Arcy, M.J. Russell, S.R. Browning and M.J. Tomlinson. The Accents of the British Isles (ABI) corpus. In *Proc. Colloque MIDL 2004*, pages 115-119, 2004.
- [5] J. Beal. English dialects in the North of England: phonology. In *A Handbook of Varieties of English*, B. Kortmann and E.W. Schneider (eds), pages 113-133, Cambridge University Press, UK, 2004.
- [6] A. Williams and P. Kerswill. Dialect levelling: change and continuity in Milton Keynes, Reading and Hull. In *Urban Voices. Accent Studies in the British Isles*, P. Foulkes and G. Docherty (eds), pages 141-162, Arnold, Londres, 1999.
- [7] J. Stuart-Smith. Scottish English: phonology. In *A Handbook of Varieties of English*, B. Kortmann and E.W. Schneider (eds), pages 47-67, Cambridge University Press, UK, 2004.
- [8] R. Penhallurick. Welsh English: phonology. In *A Handbook of Varieties of English*, B. Kortmann and E.W. Schneider (eds), pages 98-112, Cambridge University Press, UK, 2004.
- [9] M. Huckvale. ACCDIST: a Metric for Comparing Speakers' Accents. *Proc. Interspeech 2004 ICSLP*, 2004.
- [10] W.J. Barry, C.E. Hoequist and F.J. Nolan. An approach to the problem of regional accent in automatic speech recognition. *Computer Speech and Language*, 3: 355-366, 1989.
- [11] E. Ferragne and F. Pellegrino. Diphthongization as a cue for the automatic identification of British English dialects. In *Proc. 148th Meeting of the Acoustical Society of America*, 2004.
- [12] U. Altendorf and D. Watt. The dialects in the South of England: phonology. In *A Handbook of Varieties of English*, B. Kortmann and E.W. Schneider (eds), pages 178-203, Cambridge University Press, UK, 2004.