Le rythme dans les dialectes de l'anglais : une affaire d'intensité ?

Emmanuel Ferragne, François Pellegrino

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

ABSTRACT

The aim of this paper is to compare the rhythmic features of the accents of the British Isles. The first part focuses on the separability of 13 dialect classes using two-dimension representations that are common in the literature. We then introduce a new measure of speech rhythm based on intensity which enhances separability and leads to a 30,80 % correct classification score. The question of the optimal number of rhythm classes is also raised in this dialectal perspective.
Keywords: speech rhythm, phonetics, dialects, British Isles, automatic classification

1. Introduction

Les études des corrélats physiques du rythme dans la parole à partir de corpus multilingues connaissent un essor considérable depuis les années 1990 (e.g. Ramus et collègues [11], Grabe et Low [8], Dellwo et collègues [3]). Plus récemment, ces méthodes ont été employées pour l'analyse du rythme des dialectes de l'anglais (Ferragne et Pellegrino [6, 5], White et Mattys [13]).

Abercrombie [1], 222, fait allusion aux différences de quantité syllabique dans trois accents de l'anglais. À partir d'un système à 3 quantités différentes (longue, médium et brève), le mot <Peter> :
- possède le schéma long-bref dans le Yorkshire ;
- a la séquence bref-long en « Lowland Scots » ;
- peut être décrit par le schéma médium-médium en RP.

En 1982, Wells [12], 86, note que le rythme a une fonction de discrimination évidente entre les accents de l'anglais, mais qu'il reste beaucoup à faire pour qu'il soit décrit de manière satisfaisante. Il poursuit avec quelques exemples ayant trait à des différences de syllabation, de durée, de débit (le parler citadin est plus rapide que le parler rural, Wells [12], 87) et de place de l'accent. Wells [12], 362-363, rappelle également une tendance à la non réduction de certaines voyelles dans le nord, comme dans les syllabes de préfixes latins tels que *ad-, con-* et *ex-* en position prétonique. Il mentionne également le fait qu'en anglais du Pays de Galles, en syllabe finale de mot fermée, la réduction vocalique a tendance à être évitée (Wells [12], 387). Il note encore qu'en Irlande du Nord et en Écosse, la quantité, i.e. les différences phonologiques de durée, ont presque totalement disparu (si l'on omet le phénomène de « Scottish Vowel Length Rule »). Si l'on utilise une mesure de la variation de la durée des voyelles, par exemple *npviv* ou $VarcoV$ (voir Partie 3), on peut donc s'attendre à ce que, toutes choses étant égales par ailleurs (et notamment le débit), les dialectes des régions que nous venons de citer aient des valeurs plus faibles, se rapprochant ainsi (toute proportion gardée) des langues syllabiques.

White et collègues [14] ont très récemment mis au point une tâche de discrimination par des auditeurs des dialectes de l'anglais pris deux à deux à partir de stimuli de resynthèse. Les scores de classification sont très faibles et de fiabilité variable selon les paires. Les auteurs soulignent le parallèle entre les scores de classification et les mesures $VarcoV$ et $\%V$ que nous décrivons *infra*.

Dans la Partie 3, l'analyse de nos données se base sur les méthodes employées dans les articles de référence de Ramus et collègues [11], Grabe et Low [8] et White et Mattys [13]. Puis, nous évaluons une nouvelle mesure du rythme reposant sur le paramètre de l'intensité (Partie 4). Enfin, nous nous penchons sur la question du nombre de classes qu'il est possible de mettre en évidence à partir de nos paramètres (Partie 5).

2. Corpus et méthode

Le corpus Accents of the British Isles (ABI) a été enregistré en 2003 (D'Arcy et collègues [2]). Il s'agit d'une base de données payante contenant des enregistrements censés représenter 14 dialectes des Îles Britanniques. Les enregistrements ont eu lieu dans des salles assez calmes (souvent dans des bibliothèques publiques). Le signal a été capté par le biais d'un microcasque et a été numérisé directement (22 050 Hz, 16 bits). Idéalement, les locuteurs devaient avoir entre 18 et 50 ans, mais les limites réelles sont de 16 et 79 ans. Ils ont été recrutés par le biais de publicités dans les presses et radios locales. Nombre d'entre eux furent trouvés sur place à la dernière minute. Tous savaient que les enregistrements avaient pour but de mettre en évidence leur accent. Nous ne saurions trop insister sur le fait qu'aucune donnée individuelle sur l'âge, la catégorie socio-professionnelle et l'histoire linguistique des locuteurs n'est disponible, ce qui nous contraint à limiter notre étude aux deux facteurs explicites utilisables, savoir le sexe et l'origine géographique des participants.

Ces dialectes, les abréviations que nous utiliserons

Tab. 1: Dialectes du corpus ABI.

Abréviation	Dialecte	Locuteurs (H/F)
brm	Birmingham	10/10
crn	Cornwall	11/9
ean	East Anglia	9/10
eyk	East Yorkshire	13/12
gla	Glasgow	10/10
ilo	Inner London	10/11
lan	Lancashire	11/10
lvp	Liverpool	10/10
ncl	Newcastle	10/9
nwa	North Wales	10/11
roi	Republic of Ireland	10/10
shl	Scottish Highlands	11/11
sse	Standard Southern English	10/6
uls	Ulster	10/10

Fig. 1: Localités représentées dans le corpus ABI.

pour les désigner et le nombre de locuteurs sont détaillés dans la Table 1. Les lieux d'enregistrements sont localisés dans la Figure 1. Le dialecte étiqueté *ilo* a été écarté après analyse auditive par un phonéticien britannique en raison de son manque d'homogénéité. La partie du corpus utilisée ici est un passage lu d'environ 300 mots. Notre étude inclut 261 locuteurs et locutrices de 13 régions différentes.

Le signal a été segmenté automatiquement. Dans un premier temps, l'amplitude de chaque fichier a été normalisée par rapport au maximum avec le logiciel Praat. Puis les pauses, les segments vocaliques et les consonnes sont détectés grâce à un algorithme implémenté en C et Tcl/Tk (les algorithmes sont décrits dans [10]). Cette segmentation s'appuyant sur les propriétés acoustiques du signal, les frontières segmentales résultantes ne correspondent pas exactement à des entités phonologiques, mais plutôt infraphonémiques. Une fois les frontières obtenues, elles sont importées sous Praat, puis les segments adjacents de même nature (voyelles ou consonnes) sont regroupés en une seule et même entité : un intervalle vocalique ou consonantique.

3. Représentations temporelles du rythme

3.1. Paramètres

Les locuteurs sont ensuite représentés dans les 3 espaces bidimensionnels de référence :

1. le pourcentage de durée vocalique et l'écart-type de durée des intervalles consonantiques, $\%V$ et ΔC, respectivement (d'après Ramus et collègues [11]) ;
2. le « pairwise variability index » (PVI) brut pour les intervalles consonantiques (*rpvic*) et normalisé pour les intervalles vocaliques (*npviv*) donnés dans les Équations 1 et 2 (d'après, entre autres, Grabe et Low [8]) ; n est le nombre d'intervalles (consonantiques pour *rpvic* et vocaliques pour *npviv*) de la phrase, D_i est la durée de l'intervalle numéro i ;
3. le pourcentage de durée vocalique et le coefficient de variation ($VarcoV$) de durée vocalique, i.e. le rapport de l'écart-type de durée des intervalles vocaliques par la durée moyenne d'un intervalle vocalique (d'après White et Mattys [13]).

Nos mesures ne sont pas directement comparables à celles des 3 études citées puisqu'elles sont calculées à partir d'une segmentation automatique et non manuelle.

$$rpvic = \frac{\sum_{i=1}^{n-1} |D_i - D_{i+1}|}{n-1} \quad (1)$$

$$npviv = \frac{\sum_{i=1}^{n-1} |(D_i - D_{i+1})/((D_i + D_{i+1})/2))|}{n-1} \quad (2)$$

3.2. Résultats

Nous avons procédé à une analyse linéaire discriminante[1] utilisant la méthode de validation du « leave-one-out ». Les paramètres d'entrée sont chacun des 3 espaces bidimensionnels utilisés dans nos études de référence. Les taux de classification sont les suivants :
– $\%V/\Delta C$: 10,73 % ($p < 0,05$) ;
– $rpvic/npviv$: 11,88 % ($p < 0,05$) ;
– $\%V/VarcoV$: 14,56 % ($p < 0,001$).
Bien que les probabilités (test binomial) que ces taux soient dus au hasard soient faibles, ils sont nettement insuffisants pour être d'une quelconque utilité.

4. L'intensité comme paramètre du rythme

4.1. Motivation

Il est intéressant de noter que, à notre connaissance, toutes les études s'inscrivant dans la lignée de celles

[1] Avec la fonction `classify` du logiciel Matlab.

de Ramus et collègues ([11]) et de Grabe et Low ([8]) se concentrent sur le paramètre physique de la durée. Or, intuitivement, la notion de rythme n'est pas très éloignée de celle d'accent de mot et d'accent de phrase en anglais. En effet, toutes les études qui utilisent le concept de pied, ou encore celui d'intervalle entre accents, impliquent de fait que l'accent est un aspect primordial de l'impression de rythme en anglais. Et puisqu'il est avéré que l'accent de mot en anglais et la proéminence se réalisent non seulement à travers la durée, mais également l'intensité (Fry [7] pour l'accent de mot et Kochanski et collègues [9] pour la proéminence), mesurer l'intensité sous la forme d'un PVI ne semble pas incohérent pour évaluer la pertinence de ces informations. Les PVI vocaliques et consonantiques portant sur l'intensité ont été calculés à partir de la segmentation automatique de l'ensemble du passage lu de ABI. Le calcul est identique à celui décrit dans les Équations 1 et 2 si l'on remplace la durée de l'intervalle par l'intensité de cet intervalle. L'intensité moyenne en dB SPL est mesurée pour chaque intervalle avec le logiciel Praat.

4.2. Résultats

Les analyses discriminantes donnent les taux de classification correcte moyens suivants :
– npviv-I et rpvic-I : 22,05 % ;
– npvic-I, npviv-I, rpvic-I et rpviv-I : 33,84 % ;
– PVI d'intensité et de durée confondus : 30,80 %.
Un test binomial montre que ces taux de classification sont supérieurs au hasard ($p < 10^{-12}$). On remarque que lorsque les PVI d'intensité seuls sont inclus dans l'analyse, le taux de classification correcte est plus élevé que lorsque les paramètres de durée seuls sont employés.

Tab. 2: Matrice de confusion issue de la classification des dialectes à partir des paramètres d'intensité.

	brm	crn	ean	eyk	gla	lan	lvp	ncl	nwa	roi	shl	sse	uls
brm	**6**	2	2	-	-	2	-	-	1	2	-	4	1
crn	-	**8**	2	5	-	-	-	2	-	-	-	3	-
ean	1	2	**11**	3	-	-	-	-	-	-	1	1	-
eyk	-	1	7	**6**	-	-	-	-	2	-	9	-	-
gla	-	1	-	-	**6**	2	5	1	4	-	-	-	1
lan	1	1	1	-	3	**-**	3	2	4	4	1	1	-
lvp	-	-	-	-	3	-	**11**	-	-	3	-	1	2
ncl	3	3	-	-	2	-	-	**2**	4	2	-	2	1
nwa	1	5	1	1	4	1	-	1	**3**	-	-	1	3
roi	1	2	1	-	1	1	1	-	-	**1**	-	1	2
shl	-	1	-	6	-	-	-	-	-	-	**15**	-	-
sse	2	-	3	1	-	-	-	2	1	2	-	**3**	2
uls	1	-	-	-	3	-	3	3	-	-	-	2	**8**

La Table 2 donne la matrice de confusion qui rend compte des résultats de l'analyse discriminante à partir des quatre paramètres d'intensité. Les taux d'identification varient d'un dialecte à l'autre ; trois dialectes ont des taux supérieurs à 50 % : *shl* (15/22), *ean* (11/19) et *lvp* (11/20). S'il n'est pas aisé d'entrevoir les raisons pour lesquelles la classification fonctionnerait mieux pour ces dialectes, on peut néanmoins affirmer que leurs caractéristiques rythmiques (telles qu'elles sont mesurées à travers l'intensité) sont nettement distinctes des autres. À l'inverse, *lan*, avec 0 % de classification correcte semble ne pas constituer un tout suffisamment cohérent et distinct des autres dialectes dans la dimension de l'intensité.

5. Projection des dialectes dans l'espace des PVI d'intensité

Rien ne permet de penser que ces 13 dialectes soient « séparables » sur la base de leur rythme. S'il existe véritablement des classes de rythme dans les dialectes du corpus ABI, leur nombre est très certainement bien inférieur à 13. Une étude de perception permettrait d'obtenir une ébauche de réponse. À défaut d'une expérience de perception, une autre option consiste à utiliser une technique de classification avec apprentissage non-supervisé. Nous avons donc employé la méthode du *k-means clustering* pour tenter de révéler l'existence d'une structure en classes. Cette technique consiste à partitionner les données en un nombre déterminé de classes en minimisant la somme des sommes de distances d'un point au barycentre de sa classe[2]. Pour un nombre de classes allant de 2 à 13, la valeur moyenne des silhouettes, dont le calcul est donné dans l'Équation 3, est évaluée ; s_j est la valeur de silhouette du locuteur j, a_{pj} représente la distance moyenne du locuteur j aux autres locuteurs appartenant à la classe p. Si d_{qj} est la distance moyenne entre le locuteur j et tous les locuteurs appartenant à une classe $q, q \neq p$, b_{pj} est la valeur d_{qj} minimale calculée pour $q = 1...c, q \neq p$ (c étant le nombre de classes). Autrement dit, b_{pj} mesure la dissimilarité du locuteur j par rapport à la classe (autre que la sienne) la plus proche. La qualité de la partition est définie par S, la moyenne des s_j pour $j = 1...N$; dans le cas où la classe p est un singleton, $s_j = 0$. L'objectif consiste donc, si l'on cherche à déterminer le nombre optimal de classes dans une tâche non supervisée, à trouver la partition pour laquelle S est maximal.

$$s_j = \frac{b_{pj} - a_{pj}}{max\{a_{pj}, b_{pj}\}} \tag{3}$$

La valeur est bornée entre -1 et 1 ; plus la valeur est élevée, plus j est distant de la classe q la plus proche ; une valeur de 0 caractérise les locuteurs qui ne peuvent pas être clairement attribués à une classe, et une valeur proche de -1 indique que le locuteur en question n'appartient vraisemblablement pas à la bonne classe (voir Everitt et collègues [4], 104-105 et *passim*).

Les valeurs de PVI bruts et normalisés, vocaliques et consonantiques, ont été utilisées. Le nombre de classes optimal a été déterminé indépendamment pour la durée et l'intensité. Dans les deux cas que, d'après le critère de la valeur de silhouette moyenne, le nombre optimal de classe semble être 2 ($S = 0,6$). Au vu de la taille limitée des données, on peut cependant penser que ce nombre est sous-estimé. Nous avons représenté les pourcentages de classification pour la solution à trois classes (arbitrairement nommées A', B' et C' ; $S = 0,5$) par le biais d'un diagramme ternaire (Figure 2). On y relève tout d'abord l'existence du groupe *brm, ean, crn* dont la majorité des locuteurs (tous pour *brm*) appartiennent à la classe A'.

[2]La fonction `kmeans` du logiciel Matlab a été utilisée.

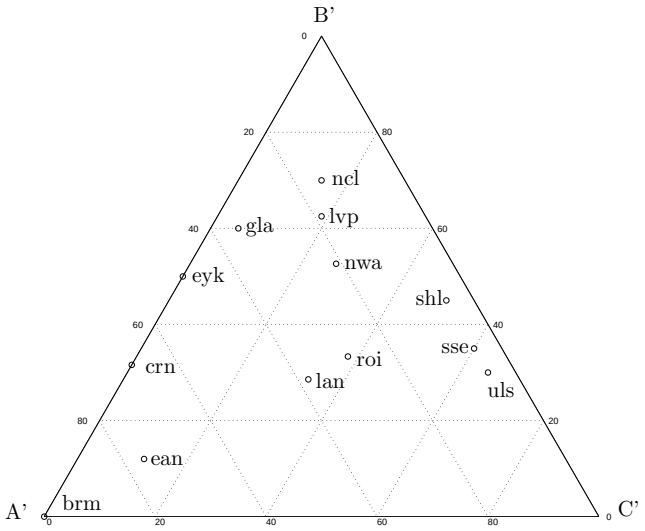

Fig. 2: Diagramme ternaire (intensité).

Ces trois dialectes forment une entité géographique cohérente : le sud de l'Angleterre (quoique, linguistiquement, *brm* soit intermédiaire entre nord et sud). La proximité des dialectes *roi* et *lan* au barycentre du triangle démontre qu'ils ont chacun une proportion approximativement identique de locuteurs dans chacune des classes. En ce qui concerne *lan*, ce résultat n'est pas surprenant puisque l'analyse discriminate (voir la Table 2) donnait 0 % de classification correcte pour ce dialecte. Les locuteurs des dialectes *ncl,lvp,gla* et *nwa* ont tendance à être regroupés dans la classe B'. La cohérence géographique ou linguistique ne peut pas véritablement justifier ce regroupement. Enfin, on voit émerger un troisième groupe rassemblant *sse* et *uls*. Pour ce dernier, s'il est certain que la durée et l'intonation systémiques diffèrent entre les deux dialectes, on peut néanmoins supposer que leurs schémas d'intensité sont proches. Everitt et collègues [4], 105, considèrent que pour $S > 0,5$, on peut véritablement parler d'une structure en classes. On peut donc inférer que les paramètres employés ici suggèrent l'existence de classes dialectales.

6. Conclusion

La transposition à la problématique des dialectes de méthodes issues d'études multilingues a permis de représenter la variation de rythme des dialectes du corpus ABI, à défaut de réellement les classer. L'introduction du PVI d'intensité constitue une véritable nouveauté ; son pouvoir discriminant s'est révélé supérieur à celui des PVI de durée. Il nous paraît donc justifié d'inclure cette mesure de la différence d'intensité moyenne entre deux intervalles vocaliques dans les études sur le rythme de l'anglais, et il serait également très intéressant de la tester dans le cadre de la classification automatique des langues. Il restera néanmoins à déterminer si d'autres mesures que l'intensité moyenne d'un intervalle (e.g. l'intensité maximale) ne seraient pas mieux adaptées.

Références

[1] David Abercrombie. Syllable quantity and enclitics in english. In David Abercrombie, D. B. Fry, P. A. D MacCarthy, N. C. Scott, and J. L. M. Trim, editors, *In honour of Daniel Jones*, pages 216–222. Longmans, Londres, 1964.

[2] S. D'Arcy, M.J Russell, S.R Browning, and M.J Tomlinson. The accents of the british isles (abi) corpus. In *MIDL*, pages 115–119, Paris, 2004.

[3] Volker Dellwo, Ingmar Steiner, Bianca Aschenberner, Jana Dankovicova, and Petra S. Wagner. Bonntempo-corpus and bonntempo-tools : A database for the study of speech rhythm and rate. In *Interspeech-ICSLP*, pages 777–780, Jeju, Corée, 2004.

[4] Brian S. Everitt, Sabine Landau, and Morven Leese. *Cluster Analysis*. Arnold, Londres, 2001.

[5] Emmanuel Ferragne and François Pellegrino. A comparative account of the suprasegmental and rhythmic features of british english dialects. In *MIDL 2004*, pages 121–126, Paris, France, 2004.

[6] Emmanuel Ferragne and François Pellegrino. Rhythm in read british english : interdialect variability. In *8th International Conference on Spoken Language Processing*, pages 1573–76, Jeju Island, Korea, 2004.

[7] D. B. Fry. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27(4) :765–768, 1955.

[8] Esther Grabe and Ee Ling Low. Durational variability in speech and the rhythm class hypothesis. In Carlos Gussenhoven and N. Warner, editors, *Papers in Laboratory Phonology 7*. CUP, Cambridge, 2002.

[9] Greg Kochanski, Esther Grabe, John Coleman, and B. Rosner. Loudness predicts prominence ; fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118 :1038–54, 2005.

[10] François Pellegrino and Régine André-Obrecht. Automatic language identification : an alternative approach to phonetic modelling. *Signal Processing*, 80 :1231–1244, 2000.

[11] Franck Ramus, Marina Nespor, and Jacques Mehler. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73 :265–292, 1999.

[12] John Christopher Wells. *Accents of English. The British Isles*, volume 2. Cambridge University Press, Cambridge, 1982.

[13] Laurence White and Sven L. Mattys. Calibrating rhythm : first language and second language studies. *Journal of Phonetics*, 35(4) :501–522, 2007.

[14] Laurence White, Sven L. Mattys, Lucy Series, and Suzi Gage. Rhythm metrics predict rhythmic discrimination. In *XVIth International Congress of Phonetic Sciences*, pages 1009–1012, Sarrebruck, 2007.