

**Université Lumière Lyon 2**

**Ecole doctorale : 3LA (Lettres, langues, linguistique et arts)**

*Centre de Recherche en Terminologie et Traduction (CRTT - EA 4162)*

**Etude phonétique des dialectes  
modernes de l'anglais des Iles  
Britanniques : vers l'identification  
automatique du dialecte**

par Emmanuel FERRAGNE

Thèse de doctorat de lexicologie et terminologie multilingues,  
traduction

sous la direction de Claude BOISSON

présentée et soutenue publiquement le 4 juin 2008

Composition du jury :

Daniel HIRST, directeur de recherche au CNRS

Claude BOISSON, professeur à l'université Lumière Lyon 2

Jacques DURAND, professeur à l'université Toulouse 2

Francis NOLAN, professeur à l'University of Cambridge

François PELLEGRINO, chargé de recherche au CNRS

# Table des matières

---

---

## Partie I Cadre de l'étude

---

---

**1**

**Introduction**

**2**

**Orientation théorique et méthodologique, et notions élémentaires**

2.1	Cadre dialectologique . . . . .	9
2.2	Variation dialectale : évolution des systèmes vocaliques et cas limites . . . . .	15
2.2.1	Les changements en chaîne . . . . .	15
2.2.2	Les convergences phonémiques . . . . .	16
2.2.3	Les scissions phonémiques . . . . .	19
2.2.4	Les quasi-convergences . . . . .	21
2.3	Cadre phonético-phonologique : variation, information et prototypes . . . . .	24
2.3.1	Variation phonético-acoustique . . . . .	26
2.3.2	Caractéristiques phonétiques des mots-test et information . . . . .	32
2.3.3	Théorie du prototype et théorie des ensembles flous . . . . .	40
2.3.4	La notion de catégorie dialectale . . . . .	52
2.4	Cadre statistique . . . . .	58
2.4.1	Les mesures de distance . . . . .	59

2.4.2	La classification hiérarchique . . . . .	61
2.4.3	L'analyse en composantes principales . . . . .	62
2.4.4	L'analyse linéaire discriminante . . . . .	62
2.4.5	« K-means clustering » et qualité d'une partition . . . . .	63
2.4.6	La régression . . . . .	64

**3**

**Situation géolinguistique et anglais standard**

3.1	L'anglais standard . . . . .	67
3.1.1	Système vocalique de l'anglais standard . . . . .	70
3.1.2	Changements récents et perspectives . . . . .	80
3.2	Caractéristiques phonético-phonologiques des dialectes des Îles Britanniques . . . . .	83
3.2.1	Partition dialectale de l'Angleterre . . . . .	85
3.2.1.1	Partition nord/sud : scission FOOT-STRUT et BATH-broadening . . . . .	85
3.2.1.2	La rhoticité . . . . .	90
3.2.1.3	L'occlusive vélaire dans <ng> . . . . .	92
3.2.1.4	Le Yod-dropping . . . . .	93
3.2.1.5	Le <i>happy</i> -tensing . . . . .	95
3.2.1.6	La voyelle de FACE . . . . .	96
3.2.1.7	La vocalisation du /l/ (L-vocalization) . . . . .	98
3.2.2	Le Pays de Galles . . . . .	99
3.2.3	L'Écosse . . . . .	102
3.2.4	L'Irlande . . . . .	104
3.2.4.1	Le sud . . . . .	106
3.2.4.2	Le nord . . . . .	108

---

---

**Partie II Description phonético-phonologique des dialectes et classification automatique**

---

---

---

**4****Introduction à l'analyse empirique**

- 4.1 Description du corpus Accents of the British Isles (ABI) . . . . . 114
- 4.2 Homogénéité et typicalité des dialectes de ABI : analyse auditive . . . 116

**5****Le rythme**

- 5.1 Préambule . . . . . 121
- 5.2 Genèse historique de la notion phonétique de rythme . . . . . 122
- 5.3 Perspectives multilingues : « stress-timing » et « syllable-timing » . . . 124
- 5.4 Analyse du rythme à partir de ABI . . . . . 136
- 5.5 Note sur le débit de parole . . . . . 156
- 5.6 Synthèse des résultats et discussion . . . . . 157

**6****Les voyelles**

- 6.1 Méthodes de représentation . . . . . 163
- 6.2 Analyse des voyelles . . . . . 170
- 6.3 Description acoustique des voyelles . . . . . 181
  - 6.3.1 L'anglais standard (*sse*) . . . . . 182
    - 6.3.1.1 Les monophthongues . . . . . 182
    - 6.3.1.2 Les diphtongues . . . . . 185
    - 6.3.1.3 Analyse des voyelles d'un locuteur type de *sse* . . . . . 188
  - 6.3.2 Birmingham (*brm*) . . . . . 206
    - 6.3.2.1 Les monophthongues . . . . . 206
    - 6.3.2.2 Les diphtongues . . . . . 209
  - 6.3.3 Cornwall (*crn*) . . . . . 215
    - 6.3.3.1 Les monophthongues . . . . . 215
    - 6.3.3.2 Les diphtongues . . . . . 218
  - 6.3.4 East Anglia (*ean*) . . . . . 221
    - 6.3.4.1 Les monophthongues . . . . . 221
    - 6.3.4.2 Les diphtongues . . . . . 226
  - 6.3.5 East Yorkshire (*eyk*) . . . . . 228
    - 6.3.5.1 Les monophthongues . . . . . 228

6.3.5.2	Les diphtongues . . . . .	231
6.3.6	Glasgow ( <i>gla</i> ) . . . . .	235
6.3.6.1	Les monophtongues . . . . .	235
6.3.6.2	Les diphtongues . . . . .	238
6.3.7	Lancashire ( <i>lan</i> ) . . . . .	242
6.3.7.1	Les monophtongues . . . . .	242
6.3.7.2	Les diphtongues . . . . .	244
6.3.8	Liverpool ( <i>lvp</i> ) . . . . .	246
6.3.8.1	Les monophtongues . . . . .	246
6.3.8.2	Les diphtongues . . . . .	251
6.3.9	Newcastle ( <i>ncl</i> ) . . . . .	254
6.3.9.1	Les monophtongues . . . . .	254
6.3.9.2	Les diphtongues . . . . .	259
6.3.10	North Wales ( <i>nwa</i> ) . . . . .	263
6.3.10.1	Les monophtongues . . . . .	263
6.3.10.2	Les diphtongues . . . . .	266
6.3.11	Republic of Ireland ( <i>roi</i> ) . . . . .	268
6.3.11.1	Les monophtongues . . . . .	268
6.3.11.2	Les diphtongues . . . . .	271
6.3.12	Scottish Highlands ( <i>shl</i> ) . . . . .	273
6.3.12.1	Les monophtongues . . . . .	273
6.3.12.2	Les diphtongues . . . . .	277
6.3.13	Ulster ( <i>uls</i> ) . . . . .	279
6.3.13.1	Les monophtongues . . . . .	279
6.3.13.2	Les diphtongues . . . . .	283
6.4	Synthèse des résultats et discussion . . . . .	285
6.4.1	Monophtongues . . . . .	285
6.4.2	Diphtongues . . . . .	303

<b>7</b>
----------

<b>Classification automatique et représentations multidimensionnelles</b>
---

7.1	Extraction des paramètres et segmentation . . . . .	310
7.2	Méthode des matrices de distances (ACCDIST) . . . . .	311
7.3	Portée descriptive et diagnostic . . . . .	321
7.4	Représentations arborées . . . . .	323

---

7.5	Détermination des voyelles discriminantes . . . . .	328
7.6	Vers une interprétation dans le cadre de la théorie des ensembles flous .	338
7.7	Synthèse des résultats et perspectives . . . . .	347

<b>8</b>
----------

<b>Conclusion générale</b>
----------------------------

<b>A</b>
----------

<b>Annexe : passage lu du corpus ABI</b>
--

<b>B</b>
----------

<b>Annexe : rythme et voisement</b>
-------------------------------------

<b>Index</b>	<b>377</b>
<b>Bibliographie</b>	<b>381</b>

**Première partie**

**Cadre de l'étude**



# Chapitre 1

## Introduction

L'étude du langage est un domaine particulièrement enclin aux changements de paradigme, à la prolifération des théories et au cloisonnement – parfois arbitraire – d'une multitude de sous-disciplines. Deux axes d'évolution épistémologique nous semblent caractériser la linguistique, au sens large. D'une part, on a pu assister au basculement d'une quête de l'invariant vers une analyse raisonnée de la variation, ce qui a permis, entre autres, le développement des études dites « variationnistes ». D'autre part, la linguistique est, à notre sens, parvenue à un stade de maturation avancé qui autorise la prise en compte de l'articulation entre constructions théoriques et phénomènes empiriques. Ainsi, la phonologie et la phonétique, disciplines naguère distinctes, présentent aujourd'hui une intersection – parfois étiquetée « phonologie de laboratoire » – qui constitue un paradigme particulièrement fécond. C'est dans ce type d'approche que s'inscrit notre travail de thèse, qui examine la structure des systèmes de sons individuels par le biais de la substance acoustique, pour ensuite regrouper les individus sur des critères de proximité phonétophonologique. Cette approche est donc, par nature, pluridisciplinaire, puisqu'elle allie la phonétique acoustique, la phonologie, la dialectologie et d'autres domaines de la linguistique dans le but de mieux cerner la complexité de la notion de dialecte sur les Îles Britanniques.

L'étude des dialectes<sup>1</sup> présente des enjeux de société singulièrement forts dans le contexte des Îles Britanniques. Les particularismes phonético-phonologiques ruraux et urbains, fréquemment érigés en stéréotypes dans les médias et les arts, reflètent l'origine géographique ou sociale d'un locuteur. C'est donc sans surprise que la dialectologie a une incidence dans des domaines aussi variés que la sociologie, la psychologie, l'ingénierie de la langue et la publicité. Au-delà de l'image vieillissante du Professeur Higgins, ce phonéticien de fiction né sous la plume de G.B. Shaw, qui s'impose la lourde tâche de faire passer une marchande de fleurs pour une duchesse en corrigeant sa prononciation, nous livrons, en guise de préliminaires, quelques exemples de l'implication directe de la problématique des dialectes dans la société.

D'après un sondage récent<sup>2</sup> mené auprès de 5000 personnes, l'acteur Sean Connery aurait l'accent le plus plaisant du Royaume-Uni. La reine, quant à elle, apparaît aussi bien dans les 10 accents préférés que dans les 10 accents les moins plaisants. Les accents les moins appréciés sont ceux de Birmingham, Liverpool et Glasgow. Notons encore que 59% des personnes interrogées souhaiteraient changer d'accent. Enfin, concernant l'accent gallois, les personnes interrogées en dehors du Pays de Galles préfèrent entendre l'anglais parlé par un Australien ou même par un Français ! La stigmatisation de certains dialectes peut avoir de lourdes conséquences. Par exemple, un article du *Irish Post* du 17 mars 2001<sup>3</sup> rapporte la mésaventure d'une jeune femme de Liverpool qui se voit refuser un emploi à Dublin en raison de son accent. Pire encore, dans une étude de Dixon et collègues ([DMC02]), un interrogatoire de police, dans lequel le suspect plaide son innocence, est présenté à des auditeurs. Cet interrogatoire, simulé, a été enregistré en deux versions ; une où le suspect adopte un accent de type RP<sup>4</sup>, une autre où il prend l'accent de Birmingham. Les auditeurs ont pour tâche d'attribuer un score de culpabilité à ce suspect.

---

<sup>1</sup>Nous emploierons ce terme pour désigner l'aspect phonétique de la prononciation d'une communauté, qu'elle soit régionale ou sociale ; il s'agira donc sans équivoque d'accents. Nous reviendrons *infra* (Section 2.1) sur le terme dialectologie.

<sup>2</sup><http://news.bbc.co.uk/1/hi/uk/4180373.stm> ; site consulté le 2 avril 2008.

<sup>3</sup>Disponible à l'adresse <http://www.phon.ucl.ac.uk/home/estuary/> ; site consulté le 10 avril 2008.

<sup>4</sup>« Received Pronunciation » ; il s'agit de l'accent britannique de prestige. Un exposé détaillé est présenté dans la Section 3.1.

---

Sans surprise, le locuteur « Brummie » est considéré comme plus coupable. Pour prendre un autre exemple, réel cette fois-ci, la police britannique, qui enquête sur une série de meurtres perpétrés par le Yorkshire Ripper à la fin des années 1970, reçoit l'enregistrement d'un homme se désignant comme l'assassin ([Ell94, Bil03]). Détail symptomatique de l'importance des dialectes sur le sol britannique, c'est à Stanley Ellis, un dialectologue renommé, que la cassette est confiée afin qu'il détermine l'origine géographique du locuteur. L'expert conclut que le locuteur est originaire de la région de Sunderland, une ville industrielle côtière du nord-est de l'Angleterre. L'enregistrement était en réalité un canular dont l'auteur n'a pas pu être identifié jusqu'en 2005 : l'ADN a en effet récemment permis de confondre celui que la presse surnomme Wearside Jack <sup>5</sup>, un certain John Humble, originaire, comme l'avait justement envisagé Ellis, de Sunderland !

En 1963 sort le premier album des Beatles, intitulé *Please Please Me* : la proportion de /r/ non pré-vocaliques effectivement réalisés atteint les 47 % ([Tru83]). En 1969 paraît le dernier album des Liverpooliens, *Abbey Road* <sup>6</sup>. Dans ce dernier, seulement 3 % de ces /r/ sont chantés. Dès l'album *Sergeant Pepper* en 1967, les mots <can't> et <half> ne sont plus exclusivement prononcés avec la voyelle /æ/ : un /a:/ plus britannique vient parfois se substituer au /æ/, rangé par Trudgill ([Tru83], 152), à l'instar de la réalisation de /r/ non pré-vocaliques, dans la catégorie des « quasi-American features ». Un début d'explication réside dans le fait qu'avant 1964, les États-Unis exerçaient une domination culturelle presque exclusive dans l'industrie du disque par le biais du rock'n'roll ; et quiconque produisait ce genre de musique se devait d'imiter le modèle américain, y compris l'accent. Après cette date, la « pop » anglaise acquiert ses lettres de noblesse et devient une culture à elle seule, ce qui permet à ses représentants de s'affranchir de la nécessité d'employer un accent américain pour vendre des disques. Quelques années plus tard apparaît le mouvement punk. La convention est alors d'adopter un accent britannique associé au bas de l'échelle sociale. Les coups de glotte se substituent à /t/ en position finale et intervocalique, les /l/ sont vocalisés, les diphtongues sont plus marquées : [æɪ] et [æʊ]

---

<sup>5</sup>[http://news.bbc.co.uk/1/hi/england/west\\_yorkshire/4825820.stm](http://news.bbc.co.uk/1/hi/england/west_yorkshire/4825820.stm) ; site consulté le 2 avril 2008.

<sup>6</sup>L'album *Let It Be* est publié en 1970, mais il a été enregistré avant *Abbey Road*.

pour <face> et <go>, respectivement ([Tru83]). L'accent, le dialecte, comme vecteur de l'identité culturelle, constitue donc un procédé publicitaire de choix.

La thématique des dialectes revêt donc un rôle crucial dans les Îles Britanniques. Notre objectif, dans ce travail, est de décrire les aspects phonétiques de 13 dialectes et de mettre au point une procédure de classification automatique des locuteurs en fonction de leur dialecte. À travers l'analyse acoustique d'enregistrements, nous explorons la prononciation de 264 anglophones dans le but de déterminer les particularités les plus diagnostiques de chaque dialecte. Si l'aspect descriptif est justifié par les points que nous venons d'énumérer, quel pourrait être l'utilité d'un système de classification automatique d'un individu selon son dialecte à partir d'un enregistrement de sa voix ? Afin de répondre à cette question, nous présentons brièvement, entre autres applications potentielles, l'intérêt de la classification du dialecte pour la reconnaissance automatique de la parole et l'indexation des voix de comédiens.

Les systèmes de reconnaissance automatique de la parole sont moins performants lorsque le dialecte de l'échantillon sur lequel s'effectue la reconnaissance ne correspond pas au dialecte d'apprentissage (i.e. le modèle). Par exemple, Yan et Vaseghi ([YV02]) ont observé, à partir d'anglais américain et britannique, une augmentation d'environ 140 à 230 % (selon la condition) du taux d'erreur lorsque le dialecte de l'échantillon et celui du modèle étaient différents.

Le site Internet <http://www.sueterryvoices.com/><sup>7</sup> est une base de données regroupant des comédiens anglophones. Immédiatement après le sexe, le deuxième critère de recherche possible est l'accent. Certes, la base est de taille modeste, mais on peut néanmoins imaginer un système d'indexation automatique des locuteurs en fonction de leur accent qui permettrait de disposer d'une base d'enregistrements de voix de comédiens bien utile dans l'industrie du spectacle et de la publicité.

Les études actuelles qui abordent les dialectes des Îles Britanniques sous l'angle de la phonétique acoustique se bornent souvent à la variation induite par des facteurs socio-

---

<sup>7</sup>Site consulté le 2 avril 2008.

---

économiques dans un seul et même centre urbain. Rassemblées dans des ouvrages collectifs (e.g. [FD99, SBK<sup>+</sup>04], etc.), ces études offrent donc en apparence l'équivalent de ce que nous proposons dans notre thèse. En réalité, si les chapitres dans [FD99] présentent l'avantage d'une description fine du lien entre structure sociale et phonétique (description rendue de fait impossible par la nature de nos données, voir Section 4.1), les méthodes et les échantillons de parole employés par les différents auteurs varient. Notre objectif est, au contraire, d'analyser la lecture d'un texte identique pour tous les locuteurs par le biais d'une méthodologie unique et reproductible (car automatique).

Dans le reste de la Partie I, nous introduisons des concepts fondamentaux : certains ont une utilité directe pour nos travaux (Sections 2.3.2 et 2.4), d'autres sont mentionnés parce qu'ils reflètent, parfois de manière indirecte, la vision épistémologique qui a accompagné nos recherches (e.g. Section 2.3.3). La Section 2.1 examine la diversité des études regroupées sous l'appellation générique de « dialectologie » et traite certains aspects de l'évolution des systèmes vocaliques. Dans la Section 2.3, les caractéristiques phonétophonologiques des voyelles sont étudiées sous l'angle de la variation acoustique, qui est ensuite éclairée par le concept d'information. Cette discussion, mise en relation avec les mots-test de notre corpus, conduit à examiner la manière dont la variation acoustique des voyelles s'inscrit dans une théorie du prototype. Enfin, la Section 2.3 s'achève sur un bref examen de ce qu'est une catégorie dialectale. Dans la Section 2.4, nous décrivons brièvement les méthodes d'analyse de données utilisées dans nos travaux. Le Chapitre 3 passe en revue les traits phonétophonologiques des variétés de l'anglais britannique. On y aborde, dans un premier temps (Section 3.1), la notion d'anglais standard dans le contexte des Îles Britanniques, accompagnée d'une revue succincte des études récentes sur le système vocalique de cette variété et sur les changements phonétiques en cours. Dans un second temps (Section 3.2), nous établissons l'inventaire des particularités phonétophonologiques les plus marquantes des aires dialectales des Îles Britanniques. L'analyse acoustique de notre corpus est détaillée dans la Partie II, qui aborde dans un premier temps la notion de rythme (Chapitre 5) pour ensuite se concentrer sur la description des

systemes vocaliques (Chapitre 6). Enfin, le Chapitre 7 est consacré à la classification automatique des dialectes et aux représentations des voyelles et des locuteurs dans un espace acoustique multidimensionnel.

# Chapitre 2

## Orientation théorique et méthodologique, et notions élémentaires

### 2.1 Cadre dialectologique

La dialectologie, au sens large, rassemble des études dont les méthodes et les objectifs sont très variés. Pour notre propos, il est possible de distinguer trois paradigmes perméables les uns aux autres :

1. la dialectologie traditionnelle, dont l'objet d'étude est principalement la dimension aréale des variations et dont les méthodes se limitent souvent à l'analyse auditive ou au questionnaire. Ce domaine s'inscrit notamment dans une optique de conservation du patrimoine ;
2. la sociolinguistique, qui étudie la variation en fonction de facteurs socio-économiques. Cette discipline constitue le paradigme dominant de nos jours, et fait appel à des techniques d'instrumentation modernes comme le spectrogramme ;
3. la dialectométrie, terme sous lequel nous regroupons les approches comportant des techniques d'analyses de données élaborées et la classification automatique.

Nous avons choisi quelques exemples d'études qui nous paraissent représentatives pour illustrer ces trois paradigmes. Le *Linguistic Atlas of England* ([OSW78]) constitue l'aboutissement d'un gigantesque projet réalisé entre 1950 et 1961 : le *Survey of English Dialects*. Durant cette période, onze enquêteurs ont sillonné l'Angleterre avec un questionnaire visant à mettre en évidence les caractéristiques des parlers de pas moins de 313 localités. Le but avoué de ce projet n'était pas d'enregistrer les variations dialectales représentatives de l'époque mais bien de garder une trace des formes de parlers traditionnels les plus conservateurs car on avait entrevu que la mobilité géographique et sociale engendrée par la guerre, accompagnée d'un effet de nivellement dû à la radio et à la télévision, allait conduire à l'extinction des particularités archaisantes de certains dialectes. Les critères retenus pour la population cible traduisent tout à fait cette volonté d'effectuer des relevés sur des traits ou des variétés moribonds : les informateurs devaient être prioritairement des hommes puisque les femmes semblent être plus disposées à assimiler les innovations linguistiques, et affichent une propension à imiter les variétés de prestige (voir [OSW78], Introduction). Ces hommes devaient être « over 60 years of age, with good mouths, teeth and hearing, and of the class of agricultural workers » ([OSW78], Introduction). En outre, dans la phase d'interprétation précédant la conception des cartes de l'atlas, Orton, dans les cas où il était confronté à des données contradictoires pour une même aire, semble avoir souvent tranché arbitrairement en faveur des formes les plus anciennes ou de celles qui déviaient le plus de l'anglais standard, ce qui n'aura pas manqué de mettre en avant des particularités qui n'étaient en réalité qu'anecdotiques.

Les accents et dialectes, notamment traditionnels, se prêtent particulièrement bien à une approche populaire – « folk linguistics », nous entendons – si bien que la collecte des données se fait souvent dans un double objectif d'étude linguistique et de divertissement familial. Cette remarque s'applique tout à fait à « The VOICES survey », projet de BBC Radio 4, qui peut être considéré comme une prolongement du *Survey of English Dialects*, et dont est tiré l'ouvrage de Elmes ([Elm05]). L'aspect scientifique de cette forme de dialectologie est assuré par des universitaires de renom, tels Stanley Ellis, l'un des investigateurs

originaux du *Survey of English Dialects*, et Clive Upton, l'un des directeurs du *Handbook of English Varieties* ([SBK<sup>+</sup>04]). Malgré cela, il apparaît comme évident que la finalité de cette enquête ressortit davantage à la préservation du patrimoine qu'à l'étude scientifique *stricto sensu* : dans les enregistrements, on retrouve, par exemple, trois coiffeuses de Truro (Cornouailles) qui livrent leurs impressions sur l'accent local, des fans de rugby à Leeds (Yorkshire) disséquant certains aspects du lexique local ; même exercice pour le pêcheur de l'île de Shetland, mais, cette fois-ci, la compréhension est ardue pour l'oreille novice. D'ailleurs, Elmes ([Elm05]) ne propose pas de transcription phonétique mais bien plutôt une tentative de rendre les particularités dialectales en orthographe standard ; le terme « eye dialect » désigne parfois ce genre d'exercice. La description des sons est d'ailleurs très impressionniste, voire fantaisiste : « pinched flat vowels » *vs* « full-throated [vowels] » ([Elm05], 32). Les enregistrements, environ 250 interviews, ont été effectués entre fin 2004 et début 2005 par une cinquantaine d'enquêteurs, et ils couvrent tout le Royaume-Uni. Le point commun entre les interviews (qui durent entre 45 minutes et 2 heures) réside dans la présentation d'un questionnaire, sur le modèle des questionnaires utilisés pour le *Survey of English Dialects*, avec des questions du type : « What do you call the toilet ? »

Voilà, donc, le type d'étude typique que nous classons dans le paradigme de la dialectologie traditionnelle. L'ouvrage de Wells ([Wel82]), quoique en partie inspiré de cette forme de dialectologie, témoigne de l'influence du tournant sociolinguistique des années 1960-70. Kerswill ([Ker03]) voit ce tournant comme « a move from the rural to the urban ». Les études actuelles s'inscrivent, pour la plupart, dans ce cadre (voir [LP97] pour un aperçu historique de cette discipline). Nous en voulons pour preuve la nature des travaux réunis dans l'ouvrage de Foulkes et Docherty ([FD99]), qui ont pour dénominateur commun la stratification sociale systématique des participants, la quantification des réalisations de ces participants (souvent limitée à des pourcentages) et, parfois, l'utilisation d'outils modernes de mesure et de visualisation du signal.

Enfin, le troisième paradigme mentionné, la dialectométrie, regroupe les études impliquant un traitement de données multidimensionnelles généralement plus élaboré, sou-

vent dans une optique de classification automatique. Ce genre d'étude regroupe, entre autres, les travaux de Barry et collègues ([BHN89]) et Huckvale ([Huc04, Huc07a, Huc07b]) pour la classification automatique dans les Îles Britanniques, Hansen et collègues ([AH02, HYHI04, AH03]) et Berkling et collègues ([BZVC98]), pour la classification automatique de dialectes anglais et américains (incluant des accents étrangers, voir aussi [BVK93]) et les travaux de Heeringa et collègues (e.g. [HN01, HG03]), ten Bosch ([tB00]), Wilhelm et Sander ([WS98]), notamment sur l'idée de distance phonétique entre dialectes.

Notre étude, contrainte dans ses objectifs par les données dont nous disposons, s'inscrit dans la lignée des travaux de dialectologie traditionnelle britannique. Certes, nous avons consulté les références provenant de la sociolinguistique variationniste, et avons rapporté les résultats pertinents émanant de ce domaine, mais la nature de nos données nous a cantonné à l'étude de la variation géographique. Si nous avons précisé « traditionnelle britannique », c'est que des habitudes codifiées distinguent les spécialistes britanniques des autres. Et s'il est un ouvrage dans lequel (la plupart de) ces conventions sont enregistrées, c'est indéniablement le *Accents of English* de John Wells ([Wel82]). Nous donnerons comme exemples, pour ne relever que ces deux-là, la terminologie et l'utilisation de la notion d'ensemble lexical. Wells ([Wel82], *passim*), même s'il n'en est pas nécessairement l'inventeur, a contribué à normaliser la dénomination de nombre de concepts, et les références actuelles en dialectologie des Îles Britanniques ([FD99], [SBK<sup>+</sup>04]) font un usage quasi-exclusif de ces termes. Concernant les ensembles lexicaux, il s'agit, à notre avis, d'une avancée méthodologique majeure qui constitue l'unique système consensuel de mots-clés dans la tradition britannique. Le principe est détaillé dans [Wel82], *passim*. Nous allons définir ce concept et tenter de prouver son utilité.

Les mots-clés désignant les ensembles lexicaux sont un moyen commode de faire référence à un groupe de mots qui ont tendance à posséder la même voyelle quel que soit l'accent, son timbre variant bien évidemment d'un dialecte à l'autre. Ils constituent ainsi une norme de description panlectale particulièrement appropriée. En règle générale, on a recours à un plus grand nombre d'ensembles lexicaux que de phonèmes pour décrire

TAB. 2.1 – Les 24 ensembles lexicaux standards, d’après Wells ([Wel82], xviii)

Mot-clé	Exemples
KIT	ship, sick, bridge, milk, myth, busy...
DRESS	step, neck, edge, shelf, friend, ready...
TRAP	tap, back, badge, scalp, hand, cancel...
LOT	stop, sock, dodge, romp, possible, quality...
STRUT	cup, suck, budge, pulse, trunk, blood...
FOOT	put, bush, full, good, look, wolf...
BATH	staff, brass, ask, dance, sample, calf...
CLOTH	cough, broth, cross, long, Boston...
NURSE	hurt, lurk, urge, burst, jerk, term...
FLEECE	creep, speak, leave, feel, key, people...
FACE	tape, cake, raid, veil, steak, day...
PALM	psalm, father, bra, spa, lager...
THOUGHT	taught, sauce, hawk, jaw, broad...
GOAT	soap, joke, home, know, so, roll...
GOOSE	loop, shoot, tomb, mute, huge, view...
PRICE	ripe, write, arrive, high, try, buy...
CHOICE	adroit, noise, join, toy, royal...
MOUTH	out, house, loud, count, crowd, cow...
NEAR	beer, sincere, fear, beard, serum...
SQUARE	care, fair, pear, where, scarce, vary...
START	far, sharp, bark, carve, farm, heart...
NORTH	for, war, short, scorch, born, warm...
FORCE	four, wore, sport, porch, borne, story...
CURE	poor, tourist, pure, plural, jury...

un dialecte car outre la diversité phonologique, la variation allophonique et l’évolution diachronique des classes de mots constituent des objets d’étude que la dialectologie ne néglige pas<sup>8</sup>.

Les ensembles lexicaux standards donnés dans la Table 2.1 ont été enrichis depuis [Wel82] au gré des besoins de chaque dialecte. Cet enrichissement permet de mieux cerner le fait que ces ensembles lexicaux ne procèdent pas de la même démarche qu’une

<sup>8</sup>Chevillet ([Che91], 32) rapporte que Gilliéron, un des pionniers de la tradition des atlas linguistiques, avait fait sienne la phrase de Grimm : « jedes Wort hat seine Geschichte ».

analyse phonologique parcimonieuse. Bien qu'ils ne soient pas mentionnés dans l'énumération dont s'inspire la Table 2.1, Wells ([Wel82], 165 *sqq.*) utilise également les ensembles *happy*, *letter* et *comma*. Or un éventuel statut phonologique de la voyelle finale de *happy* reste à démontrer (voir néanmoins la discussion détaillée dans la Section 3.2.1.5). Cependant, l'utilité d'un tel ensemble est incontestable puisque ce dernier permet d'exprimer avec concision le comportement de la voyelle inaccentuée finale d'un bon nombre des mots en <-y>, <-i>, <-ie>, <-ey>, etc. Pour prendre un autre exemple, l'ouvrage de référence de Foulkes et Docherty ([FD99]) contient un mot-clé supplémentaire : *horses*<sup>9</sup>. Phonologiquement, il s'agit d'une voyelle inaccentuée qui selon le dialecte aura la voyelle de KIT ou un schwa. Mais l'utilisation d'un nouvel ensemble lexical permet, au-delà d'une description strictement phonologique, de faire intervenir explicitement les critères morphologiques qui régissent le choix de l'une des deux voyelles citées dans les dialectes n'ayant pas encore subi ce que Wells ([Wel82], 85) nomme « Weak Vowel Merger ». Ce phénomène traduit le fait que dans certains dialectes, l'opposition /ɪ/-/ə/ permettant de distinguer traditionnellement les mots <chatted> et <chattered>, et <offices> et <officers><sup>10</sup>, a été neutralisée (ce thème sera abordé dans les Sections 3.1.1 et 3.2.1.4).

En ce qui concerne la méthodologie utilisée dans le traitement des données, notre approche s'apparente davantage aux méthodes employées dans le domaine que nous avons désigné par le terme générique de « dialectométrie ». En effet, les deux principaux objectifs de notre étude sont de classer automatiquement les locuteurs en fonction de leur dialecte et de décrire ces dialectes via des techniques de traitement de données multidimensionnelles (voir notamment le Chapitre 7). À notre connaissance, certaines méthodes utilisées ici l'ont rarement (voire jamais) été en dialectologie traditionnelle ou en sociolinguistique. Un survol sommaire de ces outils (voir *infra*, Section 2.4) permettra de mieux comprendre le reste de notre étude.

---

<sup>9</sup>Nous avons rétabli la convention typographique de Wells, qui consiste à noter la voyelle qui nous intéresse avec une petite majuscule et le reste en italiques.

<sup>10</sup>Ceci s'applique aux accents non rhotiques.

Une description des mécanismes d'évolution des systèmes vocaliques permettra d'éclairer certains aspects de la variation dialectale actuelle ; c'est l'objectif de la section suivante.

## 2.2 Variation dialectale : évolution des systèmes vocaliques et cas limites

Dans cette section, nous examinons certains facteurs expliquant les mécanismes responsables de la variation observée de nos jours dans les dialectes de l'anglais des Îles Britanniques avec une attention particulière à l'évolution des systèmes vocaliques. Nous passons en revue les trois phénomènes qui affectent l'organisation des systèmes vocaliques – changements en chaîne, convergences phonémiques et scissions phonémiques – avant d'aborder le cas limite des quasi-convergences (« near mergers »).

### 2.2.1 Les changements en chaîne

Il est nécessaire d'aborder ici la notion de changements en chaîne (« chain shifts ») car elle renforce l'idée que les inventaires vocaliques sont des systèmes (chaque voyelle n'est pas indépendante), et elle recouvre des principes qui nous permettront de mieux comprendre la variation actuelle dans les dialectes de l'anglais des Îles Britanniques. Nous nous appuyerons sur les ouvrages de Martinet ([Mar05]) et Labov ([Lab94]). L'idée de changements en chaîne se base sur les observations empiriques de la phonologie diachronique dans plusieurs langues. L'anglais a connu un tel phénomène au cours de son histoire, le Grand Changement Vocalique (GCV), et ce précédent éclaire les structures phonologiques actuelles. Labov ([Lab94], 116 *sqq.*) reconnaît trois principes généraux relatifs aux changements en chaîne :

- Principe I : les voyelles longues se ferment ;
- Principe II : les voyelles brèves s'ouvrent (non attesté dans le GCV puisque ce dernier ne concerne que les longues) ;

- Principe IIA : le noyau des diphtongues fermantes s’ouvre ;
- Principe III : les voyelles postérieures s’antériorisent (non attesté dans le GCV)<sup>11</sup>.

Un changement en chaîne élémentaire se produit lorsqu’un phonème A quitte sa position initiale dans le trapèze vocalique, et qu’un phonème B vient occuper cette position. Ceci ne nous dit rien sur le sens de la causalité ; il se peut que ce soit A qui ait amorcé le mouvement, auquel cas on parlera de chaîne de traction ; si c’est B, on parlera de chaîne de propulsion. Lorsque choisir l’une de ces deux options a un sens, la décision sera guidée par les principes énoncés plus haut ainsi que par des critères sociolinguistiques.

Il apparaît également que l’évolution des systèmes vocaliques préserve une certaine symétrie entre les voyelles antérieures et postérieures. Ceci a été vrai diachroniquement : les voyelles de PRICE et MOUTH étaient les monophthongues longues les plus fermés du système moyen-anglais ; le processus de fermeture des longues (ici « remplacé » par une diphtongaison à cause de l’impossibilité d’atteindre un degré de fermeture supérieur) suivi de l’ouverture des noyaux des diphtongues fermantes a affecté ces voyelles de la même manière. En synchronie, il est intéressant de noter une symétrie sur le plan de la réalisation pour des voyelles d’aperture égale ; par exemple, dans ce que nous appelons plus loin « anglais standard », les deux voyelles actuelles les plus fermées (FLEECE et GOOSE) sont légèrement diphtonguées et pourraient, incidemment, emprunter le même chemin que PRICE et MOUTH cinq siècles plus tôt.

### 2.2.2 Les convergences phonémiques

Les convergences phonémiques (« phonemic mergers ») s’opposent au phénomène de changements en chaîne en cela que, schématiquement, le mouvement d’un phonème A vers le phonème B ne provoque pas un mouvement de ce dernier, ce qui conduit A et B à être confondus au final (le cas décrit est en réalité un des trois mécanismes reconnus dans les convergences). Labov précise que les convergences phonémiques sont bien plus courantes

---

<sup>11</sup>Mais, nous le verrons plus loin, ce phénomène est attesté, notamment en anglais standard, pour les ensembles FOOT et GOOSE.

dans l'histoire des langues que les changements en chaîne ([Lab94], 310).

D'après Labov ([Lab94], 310-348), les convergences phonémiques obéissent à deux règles : le principe de Garde<sup>12</sup> et le principe de Herzog<sup>13</sup>. Le premier pose que les convergences phonémiques sont irréversibles ou, comme le formule Labov ([Lab94], 311) : « Mergers are irreversible by linguistic means ». Le second complète le précédent : l'expansion des convergences se fait au détriment des distinctions ([Lab94], 311 : « Mergers expand at the expense of distinctions »).

La justification du principe de Garde est très intuitive : une fois que deux classes de mots, deux ensembles lexicaux, ont fusionné dans une langue ou un dialecte, le locuteur n'a dans sa phonologie plus qu'une seule catégorie et ne peut normalement pas reconstituer les deux classes historiques. C'est par exemple le cas chez les locuteurs d'anglais d'Écosse qui ne présentent pas la distinction FOOT-GOOSE. Si l'un d'entre eux souhaite, mû par une quelconque pression sociolinguistique, réapprendre ces deux catégories qui continuent d'être distinctes dans la majorité des dialectes des Îles Britanniques, il devra réapprendre la prononciation de chaque item. L'orthographe n'étant pas un guide fiable pour ces deux classes, il courra en outre le risque de l'hypercorrection<sup>14</sup>. Labov [Lab94], 312, ajoute que ce principe bénéficie d'une validité empirique certaine :

at no known time in the history of languages has such a reversal been accomplished by enough individual speakers to restore two original word classes...

Il existe cependant des exemples de cas limites où les deux classes de mots apparemment impliquées dans une convergence ont été reconstituées. Labov ([Lab94], 306-309) et Wells ([Wel82], 208-210) expliquent qu'une convergence (au moins partielle) entre les ensembles PRICE et CHOICE<sup>15</sup> est attestée dès le 15<sup>e</sup> siècle. Ce phénomène semble avoir atteint son paroxysme au 17<sup>e</sup> siècle. À la fin du 18<sup>e</sup>, il s'agissait d'un trait hautement stigmatisé. De nos jours, la plupart des accents ont rétabli la distinction. Wells ([Wel82],

---

<sup>12</sup>Du nom d'un universitaire slavisant français, Paul Garde.

<sup>13</sup>Marvin Herzog est un collègue de Labov, spécialiste du yiddish.

<sup>14</sup>Par exemple, la forme graphique ne permet pas de prédire qu'en anglais standard <wood> ne rime pas avec <mood> ; statistiquement, /u:/ est plus courant pour le digraphe <oo> ([Jon03], 381).

<sup>15</sup>Notre formulation suppose un certain raccourci : ces ensembles lexicaux sont ceux de l'anglais actuel ; or la convergence que nous rapportons ne semble concerner qu'un sous-ensemble historique de CHOICE.

209-210) attribue ceci à l'influence de RP (sans plus de précision) et à la pression exercée par l'orthographe. Ces deux facteurs sont probablement de deux types bien différents ; si l'orthographe constitue très vraisemblablement une condition nécessaire pour le rétablissement d'une distinction perdue après convergence, l'hypothèse de pressions sociolinguistiques causant le processus paraît la moins extravagante. Et si des pressions sociales ont pu initier ce processus, c'est qu'il n'était pas achevé (voire pas initié) dans les variétés qui ont fini par imposer leur norme, d'où une certaine réserve quant à l'ampleur géographique et sociale de cette convergence, et son degré de complétude. En étudiant le phénomène de plus près et en suivant l'un de ses principes méthodologiques récurrents (étudier le présent pour comprendre le passé), Labov ([Lab94], 371-384) observe la convergence PRICE-CHOICE dans l'Essex. Il conclut que ces deux voyelles n'ont probablement pas connu une totale convergence au 18<sup>e</sup>, mais une quasi-convergence (« near merger », phénomène déconcertant pour le linguiste structuraliste ; voir ci-dessous 2.2.4). En tout état de cause, pour Labov, s'il n'y a pas eu convergence, c'est donc qu'on ne peut pas véritablement parler de restauration de deux classes après convergence ; le principe de Garde reste donc intact.

D'après Labov ([Lab94], 321 *sqq.*), il existe trois mécanismes distincts pouvant aboutir à une convergence phonémique :

- la convergence par approximation ; les réalisations de deux phonèmes se rapprochent au point de finir par se confondre. Le phonème unique résultant de ce processus présente soit une qualité phonétique intermédiaire entre les deux timbres originaux, soit une réalisation correspondant à celle de l'un des deux phonèmes de départ. Ce genre de convergence peut se mettre en place en trois ou quatre générations ;
- la convergence par transfert ; les mots d'une classe rejoignent progressivement une autre classe. C'est le processus le plus long ;
- la convergence par expansion ; les réalisations de deux phonèmes se dilatent dans l'espace vocalique au point de finir par se chevaucher. L'espace phonétique occupé par les réalisations du nouveau phonème englobe les deux espaces précédemment

occupés par les réalisations des deux phonèmes. C'est le type de convergence le plus rapide ; il peut s'effectuer en une seule génération.

### 2.2.3 Les scissions phonémiques

Les scissions phonémiques (« phonemic splits ») sont l'inverse des convergences : la naissance d'une nouvelle distinction phonémique dans le système due à la scission d'une seule catégorie phonémique préexistante. Abordons dans un premier temps l'influence des emprunts pour ensuite détailler le cas où une classe de mots en vient à être scindée sans règle apparente.

Les emprunts à d'autres langues peuvent enrichir un inventaire phonémique de deux manières différentes : soit en apportant un nouveau phonème, comme cela semble être le cas pour les mots de l'ensemble CHOICE ([Wel82], 150), soit en permettant la phonologisation de ce qui n'était qu'une variante allophonique dans la langue qui emprunte. Ce dernier point est illustré par la phonologisation de l'opposition de voisement pour certaines fricatives en moyen-anglais. Hormis dans les régions du sud-ouest de l'Angleterre, le vieil-anglais ne possède pas de fricatives voisées à l'initiale d'un mot ; en revanche, celles-ci se sont voisées dans cette période lorsqu'elles apparaissaient entre deux segments voisés, d'où les doublets actuels : <wife>/<wives><sup>16</sup>, <bath>/<bathe>, etc ([Sté97], 18 *sqq.*). Il s'agit donc ici de variation allophonique conditionnée par le contexte. En conjonction avec d'autres facteurs (dont le voisement dû à l'affaiblissement des fricatives à l'initiale de mots grammaticaux, e.g. <this>, <thy>, etc.), les emprunts au français ont permis la présence de fricatives voisées à l'initiale, et ont ainsi contribué à la formation de paires minimales (e.g. <fine>/<vine>).

Ce que Labov ([Lab94], 333 *sqq.*) nomme « lexical split » implique un processus de scission d'un ensemble lexical dont les raisons sont multiples et moins intuitives que l'influence des emprunts. L'exemple du « broad a », ce phénomène historique typiquement

---

<sup>16</sup>Le <-e> du singulier n'est qu'un diacritique conventionnel qui signale la longueur de la voyelle, et non une relique désinentielle qui aurait pu favoriser le voisement de la fricative ; en vieil-anglais : <wif> [wi:f]/<wifas> [wi:vas] ([Sté97], 18).

TAB. 2.2 – Illustration de la difficulté du choix de l’affectation d’un item à l’ensemble BATH.

TRAP : /æ/	BATH : /ɑː/
classic	class
gas	grass
ample	sample
land	demand
math	bath
plastic	plaster
cancer	answer

britannique qui consiste en l’allongement et la postériorisation de la voyelle dans les mots de l’ensemble BATH illustre la complexité des facteurs qui interviennent. Certains critères phonologiques (« broad a » notamment devant /f/, /θ/, /s/ et /n/ + obstruante sourde), ou morphémiques (<chancy> est un dérivé; le dérivant remplit une condition phonologique favorable au « broad a », alors que <fancy>, qui n’est pas un dérivé, a /æ/) permettent à qui voudrait apprendre cette distinction de connaître certaines tendances. Mais, comme le montre la Table 2.2 les irrégularités sont nombreuses, d’où, une fois de plus, le risque d’hypercorrection.

Il apparaît donc qu’acquérir ce type de distinction nécessite souvent l’apprentissage un à un des items susceptibles d’appartenir à la nouvelle classe. Pour prendre un autre exemple, les dialectes du nord de l’Angleterre, qui ont la même voyelle pour FOOT et STRUT jouissent de relativement peu de prestige (voir Section 3.2.1.1). Il est donc envisageable qu’un locuteur natif d’un de ces dialectes soit amené à apprendre cette distinction au contact prolongé de locuteurs méridionaux ou de locuteurs de classes supérieures. Or, si l’orthographe constitue une règle probabiliste assez fiable (la graphie <u> étant typique de STRUT), il risque cependant de produire certaines formes hypercorrectes s’il ne se fie qu’à celle-ci. Ainsi l’orthographe le conduira à inférer à raison que <but> appartient à STRUT, mais que <butcher> est dans le même cas à tort. S’il prononce <Hull> avec /ʌ/, il devra se garder de généraliser cette prononciation à <full>, <pull>, etc. Et s’il constate que les singularités orthographiques du genre de <love> ou <come><sup>17</sup> ont la voyelle de

---

<sup>17</sup>Historiquement, ces mots avaient comme voyelle un <u> dans la graphie; s’il est conservé dans la prononciation, il a en revanche été remplacé par un <o> dans l’écriture cursive pour faciliter sa lisibilité,

STRUT, il devra s'abstenir d'en déduire que les étrangetés orthographiques que sont la première voyelle de <woman> et de <bosom> appartiennent au même ensemble que <love> et <come>. À ce propos, Gerry Knowles, pourtant phonéticien (mais originaire du nord) écrivait en 1978 ([Kno78], 86) :

There are just a few words like *putty* and *nasty* which baffle me; I cannot recall whether RP has /pʊtɪ/ or /pʌtɪ/, /næstɪ/ or /nɑːstɪ/, and I have to look them up in a pronouncing dictionary.  
(Soulignement ajouté)

## 2.2.4 Les quasi-convergences

Le phénomène des quasi-convergences (« near mergers ») est un cas délicat car il illustre les limites d'une phonologie structuraliste trop conservatrice.

Labov ([Lab94], 371-384) aborde plus en détail la convergence PRICE-CHOICE. Au début des années 1970, il se rend à Tillingham<sup>18</sup> pour mener une étude sur trois informateurs âgés de 69 à 87 ans. Il parvient à éliciter les paires <voice>/<vice> et <loin>/<line>. À la question de savoir si les deux membres de chaque paire sont identiques ou différents, deux informateurs répondent qu'ils sont identiques, le troisième, différents. Sur le plan de la production, les trois informateurs ont des valeurs sur un plan F1/F2 qui, quoique assez semblables pour PRICE et CHOICE, font apparaître deux groupes plutôt distincts. En faisant écouter aux trois informateurs les hypothétiques paires minimales produites par l'un d'entre eux, aucun des trois ne parvient à distinguer les membres de chaque paire. La discussion qui suit entre l'investigateur et deux des sujets amène ces derniers à conclure qu'une différence existe, l'un d'entre eux tentant de l'exagérer (probablement inspiré par ce qu'il connaît des autres dialectes ou par la graphie), et l'autre en fournissant une imitation très approximative, le contraste étant surtout rendu par le contexte sémantique de ses commentaires. Les conclusions de cette étude sont appuyées par les résultats d'autres études résumées par Labov ([Lab94], 357-370).

---

et l'habitude a été conservé malgré l'avènement de l'imprimerie pour certains mots fréquents ([Sté97], 42).

<sup>18</sup>Petit village de l'Essex où les enquêteurs du *Survey of English Dialects* avaient diagnostiqué la persistance de la convergence PRICE-CHOICE.

La linguistique pose généralement une symétrie entre production et perception : quand il parle, le locuteur est aussi auditeur, et quand il écoute, l'auditeur est aussi locuteur ([Lab94], 352). Or le cas des quasi-convergences fait apparaître qu'il est possible pour un seul et même locuteur de produire des différences phonétiques de façon cohérente sans les percevoir. Se pourrait-il, donc, qu'une paire minimale continue d'exister sur le plan phonétique tout en ayant perdu sa valeur phonologique pour le locuteur qui la produit ? De plus, se pourrait-il que ce phénomène infraliminal ait une quelconque influence sur l'évolution d'un système ?

Le problème des quasi-convergences peut être mis en parallèle avec certains aspects de phonologie développementale, notamment les « covert contrasts ». On sait que le nourrisson est capable de distinguer très tôt des contrastes phonologiques dans une langue étrangère qui sont implémentés phonétiquement de manière différente de sa langue maternelle ([Jus96]). Il lui est en outre possible de distinguer des contrastes inexistant dans sa langue maternelle, comme c'est le cas de ces nourrissons canadiens (« anglophones ») âgés de 1 à 4 mois qui discriminent /pa/ et /pã/ en français dans une étude rapportée par [Jus96], 334. Le nourrisson est donc un incroyable phonéticien, et s'il est doué d'une perception avec une granularité aussi fine à ce stade immature de l'acquisition de sa phonologie, ceci lui confère probablement la capacité de produire de la variation phonétique (pertinente ou non) imperceptible pour un adulte. Cette remarque est corroborée par un certain nombre d'études, dont celle de Scobbie *et al.* ([SGHF00]). La plupart des études sur l'acquisition de la phonologie se basent sur des transcriptions phonétiques d'enregistrements. On sait par ailleurs que, dans ce domaine, la comparaison de transcriptions d'un même passage par plusieurs transcrip-teurs expérimentés montre des différences très importantes ([SGHF00], 194-196), d'où une fiabilité douteuse. Ces transcriptions font croire que beaucoup d'oppositions présentes chez l'adulte sont neutralisées chez l'enfant. Or les études citées par [SGHF00] prouvent que l'enfant dans la phase d'acquisition de la phonologie parvient à contraster en production de façon cohérente certains phonèmes de sa langue, le contraste étant imperceptible pour un adulte. Par exemple, Scobbie *et al.* ([SGHF00]) montrent

dans une étude sur l'acquisition du trait de voisement, que leur sujet produit des valeurs de VOT et de pentes spectrales différenciées selon que la consonne articulée est phonologiquement voisée ou non. Ces mesures parviennent, là où l'oreille adulte échoue, à prouver l'existence de catégories phonologiques bien établies, quoique opposées en surface par des indices acoustiques qu'on pourrait qualifier de subliminaux pour un auditeur adulte. Ces faits appellent un constat et soulèvent une question. D'abord, l'étude de l'acquisition de la phonologie sans l'appui de la phonétique expérimentale est une entreprise hasardeuse. Ensuite, le fait que l'enfant en phase d'acquisition de sa phonologie puisse produire et, peut-être, percevoir des contrastes que le système de l'adulte ne reconnaît pas, ou plus du tout, peut-il jouer un rôle dans l'évolution d'une langue ou d'un dialecte ?

Dans le même ordre d'idées, Scobbie ([Sco07]) étudie le phénomène de dérhoticisation chez deux adolescents écossais issus de la classe ouvrière. Ce phénomène fait référence à la disparition (sur le plan de la perception) du /r/ en position de coda. L'étude compare dans un premier temps un spectrogramme de <rain>, où le /r/ initial se traduit par un mouvement rapide vers le haut de F2 et F3 sur un peu plus de 100 ms, avec un spectrogramme de <car> où l'on observe, là où le /r/ est attendu, un mouvement extrêmement réduit des deux formants, avec une intensité bien moindre que celle de la voyelle, et sur une durée particulièrement faible ; en d'autres termes, rien sur le spectrogramme n'indique la présence d'une approximante alvéolaire ou post-alvéolaire. En utilisant une technique d'imagerie à ultrasons<sup>19</sup> (« Ultrasound Tongue Imaging »), Scobbie met en évidence, à la fin de <car>, un mouvement de la lame de langue en direction de la région post-alvéolaire tout à fait semblable à celui que le locuteur produit à l'onset de <rain>. Son absence sur le plan de la perception semble provenir du fait que ce mouvement de la langue intervient plus tard que chez les locuteurs qui préservent une rhoticité perçue. Il apparaît donc que, chez ce locuteur, le /r/ imperceptible en coda soit néanmoins présent acoustiquement sous une forme vestigiale.

Ce rapide tour d'horizon des phénomènes « évolutionnistes » avait pour but de carac-

---

<sup>19</sup>Proche, quant au principe et à l'image obtenue, d'une échographie classique.

tériser la variation systémique dans les dialectes. Dans la section suivante, nous abordons la question de la variation acoustique.

## 2.3 Cadre phonético-phonologique : variation, information et prototypes

L'orientation phonético-phonologique de nos travaux est en grande partie déterminée par la nature de nos données. En effet, la taille importante du corpus utilisé (voir Section 4.1) justifie une étude exclusivement acoustique ; les aspects de perception constituant un développement ultérieur potentiel. Le lien entre phonétique et phonologie ou, autrement dit, entre des paramètres physiques variant dans un espace continu et des catégories linguistiques, nécessite une approche qui, au moins dans un premier temps, s'éloigne de la représentation classique structuraliste qui prend la forme d'une matrice de traits binaires. Pour Chomsky et Halle, la projection de la substance phonétique (par nature gradiente) dans un espace phonologique binaire (ou inversement) n'a rien de contradictoire ([CH68], 297) :

In view of the fact that phonological features are classificatory devices, they are binary, [...] for the natural way of indicating whether or not an item belongs to a particular category is by means of binary features. This does not mean that the phonetic features into which the phonological features are mapped must also be binary<sup>20</sup>.

Une revue des développements postérieurs à Chomsky et Halle sur l'idée de correspondance entre les représentations phonétique et phonologique est présentée dans les travaux de Chitoran et Cohn ([CCis]). Les auteurs abordent dans un premier temps la représentation modulaire classique, qui fait co-exister une phonologie strictement catégorielle avec une phonétique gradiente. Elles traitent ensuite l'approche modulaire alternative, qui plaide en faveur d'une phonétique discrète et d'une phonologie plus gradiente. Comme nous le verrons, notamment lorsque nous ferons allusion à la théorie quantique de Stevens (Section 2.3.3), il existe dans la substance phonétique – supposée varier de

---

<sup>20</sup>Il nous apparaît néanmoins que, contrairement à ce que semblent penser les auteurs, une logique binaire – quoique autorisant des représentations parcimonieuses – n'est *a priori* pas plus « naturelle » qu'une logique plurivalente comme, par exemple, celle que nous abordons *infra* Section 2.3.3.

manière continu – des zones de stabilité qui imposent une certaine discontinuité à cette substance<sup>21</sup>. Un troisième paradigme propose une approche unifiée de la phonétique et de la phonologie<sup>22</sup>. En effet, la frontière entre les deux n'est pas toujours évidente, comme le montrent Scobbie et Stuart-Smith ([SSS06]) dans une étude sur les contrastes « quasi-phonémiques ». Les auteurs donnent l'exemple de la *Scottish Vowel Length Rule*, qui aboutit à des paires comme <crude> [k<sup>h</sup>rʉd] vs <crewed> [k<sup>h</sup>rʉ:d], <side> [sʌɪd] vs <sighed> [sʌɪd], etc<sup>23</sup>. Chaque paire pourrait être considérée comme une paire minimale puisque une différence phonétique engendre une différence de sens. Or l'alternance de la voyelle – au moins pour les deux paires citées – est entièrement prévisible par la morphologie. Ces phénomènes sont en outre complexifiés par le fait que seules certaines voyelles sont soumises à cette règle d'allongement. Les auteurs concluent que ces phénomènes sont pertinents pour une analyse phonologique (même conservatrice), et qu'il est nécessaire, dans le cadre d'une phonétique/phonologie modulaire, de pouvoir les classer soit comme entités phonologiques, soit comme formes de surfaces dérivables d'une représentation sous-jacente. Mais, pour ce faire, il faut laisser à la phonologie la flexibilité de représenter plusieurs degrés de phonologisation afin d'intégrer dans l'analyse des cas limites comme la *Scottish Vowel Length Rule*.

Dans cette section, après avoir passé en revue les facteurs engendrant la variation des quantités physiques que nous étudions, l'approche des catégories que nous avons employée est détaillée. Cette discussion est indispensable compte tenu du fait que dans le Chapitre 6.2, nous inférons le système phonologique des dialectes à partir de mesures acoustiques : la correspondance entre la substance et la forme doit donc être rendue explicite pour justifier nos inférences.

---

<sup>21</sup>La représentation de l'intonation par le biais d'un « prosogramme » (voir par exemple Patel et collègues [PIR06]) opère une discrétisation de la courbe de F0 qui illustre un type de phonétique plus catégoriel.

<sup>22</sup>Le succès de la « phonologie de laboratoire » illustre bien l'importance de ce nouveau paradigme (voir, à ce sujet, [DL02], 38-40).

<sup>23</sup>Ce phénomène est en réalité beaucoup plus complexe.

### 2.3.1 Variation phonético-acoustique

Le thème de la variation, qui est récurrent en phonétique (notamment acoustique), semble exercer une certaine fascination. Il recouvre plusieurs concepts que nous allons tenter de démêler, tout en nous efforçant de démontrer que la variation phonétique n'est pas plus imprévisible (ou mystérieuse) que d'autres types de variations auxquelles l'être humain est confronté quotidiennement. Les facteurs de variabilité acoustique peuvent être regroupés en trois catégories :

1. la variation déterminée par des propriétés physiologiques ;
2. la variation déterminée par le contenu informationnel du signal de parole ;
3. la variation linguistique ; celle qui persiste une fois que les deux autres (ci-dessus) ont été écartés.

On sait depuis les premières études de grande ampleur en phonétique acoustique qu'une voyelle, telle qu'elle est mesurée par le biais des proéminences spectrales appelées formants, varie d'un locuteur à l'autre, entre hommes, femmes et enfants, et également chez un seul et même locuteur répétant plusieurs fois la même voyelle (e.g. [PB52]). Concernant les deux premiers types, cette variation est, toutes choses (dialecte, style, etc.) égales par ailleurs, le corollaire de différences physiologiques permanentes. En effet, les hommes ayant, par exemple, des conduits vocaux en moyenne plus longs que ceux des femmes, les formants de ces dernières sont plus élevés et plus espacés (voir e.g. [KDL96] ; d'autres variables physiologiques sont mentionnées dans [Cal89], 86-87). Malgré le fait que les valeurs acoustiques d'un /i/ en anglais américain soient en moyenne, pour les deux premiers formants, de 270 et 2300 Hz pour les hommes, et de 300 et 2800 Hz pour une femme ([KDL96], 205), c'est bien à une seule et même catégorie phonologique que l'auditeur associe ce son. Il n'y a aucune raison *a priori* de penser que cette variation ne suive pas une loi normale. En revanche, ceci n'implique pas que les différences de valeurs acoustiques observables entre les hommes et les femmes soient exclusivement une conséquence de la variation physiologique ; en comparant les valeurs formantiques des trois premiers formants vocaliques entre hommes et femmes dans 26 langues, Johnson ([Joh05], 379-381) démontre, au vu de la variation importante d'une langue à l'autre, que des aspects culturels ou sociophonétiques entrent en ligne de compte. Ce double constat nous poussera à dissocier les hommes et les femmes dans l'analyse de la Section 6.2. En ce qui concerne la

variation de valeurs formantiques pour un seul et même locuteur produisant  $n$  répétitions d'une voyelle, tous les autres facteurs (nous les détaillons *infra*) étant contrôlés, elle est imputable, en grande partie, aux limites inhérentes à la précision du contrôle moteur des articulateurs et à la tolérance du système auditif et de l'organisation mentale des catégories phonémiques vis-à-vis de cette imprécision et de la distance objective entre deux sons. À ce sujet, Broad ([Bro76]) réexamine les productions de plusieurs voyelles d'un même locuteur issues d'études antérieures et constate une variation intra-catégorielle (suivant la loi normale) très faible et une distance inter-catégorielle importante ([Bro76], 403) :

all the studies demonstrate a remarkably tight clustering of the formant frequencies for repeated items. In fact, the standard deviations are of the same order as the human auditory difference limen for formant frequencies. . .

Un argument censé prouver la variabilité atypique de la parole consiste à remarquer la grande dispersion et le chevauchement inter-catégoriel dans l'espace des deux premiers formants mesurés en un point de l'état stable d'une voyelle. Or d'autres paramètres moins variables ont été mis en évidence, par exemple, dans l'étude de Verbrugge et Rakerd ([VR86]). Les sujets ont pour tâche d'identifier des voyelles présentées dans un contexte consonantique /b\_b/<sup>24</sup>. Les stimuli originaux sont des monosyllabes produits en isolation par un homme et une femme. Les sujets sont soumis à 5 conditions :

- la syllabe entière, c'est la condition de contrôle ;
- les 20 % du début et de la fin de la syllabe, avec un silence (correspondant à la partie stable de la voyelle) entre les deux ;
- idem que la condition précédente, excepté que les premiers 20 % proviennent des réalisations du locuteur et les derniers, de la locutrice, ou vice-versa ; c'est la condition dite « hybride » ;
- uniquement les 20 % à l'initiale ;
- uniquement les 20 % en position finale.

Contrairement aux résultats d'études antérieures qu'ils citent, les auteurs notent une légère dégradation de l'identification entre les deux premières conditions. En revanche, les

---

<sup>24</sup>Ce type d'expérience est regroupé sous l'étiquette « silent centers » ; voir Jenkins et collègues ([JST99]).

taux d'erreurs dans la deuxième et la troisième condition (23,1 et 27,4 % respectivement) ne sont pas significativement différents. Les deux dernières conditions ont des taux d'erreur de 56,4 et 73,8 %. Cette étude confirme ce qui avait déjà été établi : la conjonction des informations dynamiques présentes au début et à la fin d'une syllabe CVC permettent une bonne identification de la voyelle nucléaire lorsque la cible de celle-ci est remplacée par un silence. La présentation de la partie initiale ou finale seule de cette syllabe engendre des taux d'erreur élevés, ce qui suggère que ce sont bien les deux transitions à la fois qui sont pertinentes dans l'identification d'une voyelle. Le résultat le plus inattendu réside dans le fait que l'identification est aussi bonne dans la condition hybride que dans la deuxième condition. Ce dernier point permet d'émettre l'hypothèse que l'information dynamique aux marges d'une voyelle dans un contexte CVC varie peu par rapport à la cible vocalique entre un homme et une femme. Les auteurs réaffirment<sup>25</sup> le rôle des transitions comme indices acoustiques propres et rejettent l'idée selon laquelle ces transitions ne feraient que permettre à l'auditeur une extrapolation visant à prédire le spectre de la partie stable (la cible) de la voyelle (conclusion identique dans une étude de Jenkins et collègues ([JST99])). Dans le cadre de notre brève discussion sur l'apparent paradoxe de l'invariance, cette étude démontre que l'information acoustique classiquement mesurée au niveau de la cible vocalique, et dont l'extrême variabilité sert d'argument appuyant l'impossibilité de résoudre ce paradoxe, peut être contrebalancée par des indices dynamiques relativement insensibles aux variations physiologiques entre locuteurs<sup>26</sup>.

Voyons maintenant la contribution des limites du système auditif au degré de liberté articulaire (et donc acoustique). Le système auditif est décrit en détails dans [Gre96]. L'exemple de l'échelle des fréquences illustre comment celui-ci autorise la variation acoustique. Le fonctionnement de la cochlée engendre des limites de résolution fréquentielle variables. Autrement dit, la fréquence objective et la fréquence telle qu'elle est restituée

---

<sup>25</sup>La complémentarité des valeurs d'état stable et de pentes formantiques dans l'identification des voyelles avait déjà été démontré par Lindblom et Studdert-Kennedy ([LSK67]) entre autres.

<sup>26</sup>Al-Tamimi ([AT07]) aborde la question des pentes formantiques dans les dialectes de l'arabe et en français ; il ré-affirme l'importance de la conjonction des caractéristiques spectrales statiques et dynamiques dans la production et la perception de la parole.

au niveau de la cochlée n'entretiennent pas une relation linéaire. Plus précisément, plus la fréquence est élevée, plus l'écart de fréquence objective entre deux sons différents doit être élevé pour être interprété comme identique par le système auditif. Ceci a pour conséquence directe de favoriser la variation acoustique objective au fur et à mesure que l'on monte dans l'échelle des fréquences. Cette variation est donc une nouvelle fois imposée par des caractéristiques physiologiques ; elle ne présente donc que peu d'intérêt en tant qu'objet d'étude *per se* pour la description phonétique, d'autant plus que des échelles psycho-acoustiques ont été mises au point pour tenter de la maîtriser (voir *infra*). Corrélativement à ces phénomènes, et en rapport direct avec notre étude, la question des seuils différentiels de perception des formants vocaliques doit être abordée. Il s'agit ici de déterminer la déviation nécessaire dans le plan F1/F2 pour qu'un stimulus vocalique soit jugé comme différent d'un stimulus de référence. Les formants étant étudiés séparément, le seuil différentiel de perception est la différence entre la fréquence formantique du stimulus de référence et celle du stimulus le plus proche constituant néanmoins une différence perceptible.

Flanagan ([Fla55]) obtient des seuils variant de 12 à 27 Hz pour F1 (entre 300 et 700 Hz), et de 20 à 90 Hz pour F2 (1000 à 2000 Hz). Il observe en outre une asymétrie entre les valeurs de seuil inférieures et celles supérieures au stimulus de référence. Les études suivantes (rapportées dans [KPW94], 486) font apparaître des seuils allant, en moyenne, de 2,8 à 9,6 % de la valeur F1, et de 2 à 8,9 % de la valeur de F2. L'étude de Kewley-Port et Watson ([KPW94]) ne rapporte pas une asymétrie marquée ; leur seuil, peu variable d'un timbre à l'autre, est de 14,5 Hz (à peu près constant) sur F1 (<800 Hz), et augmente linéairement de 10 Hz tous les 1000 Hz sur F2. Ces études, menées dans des conditions de laboratoire, avaient pour but de détecter un changement perceptible de qualité du timbre vocalique et non un changement de catégorie phonémique. Il est donc évident que :

- la différence de fréquence opposant les réalisations types de deux phonèmes distincts est largement supérieure à ces seuils ; plus généralement, les études sur la perception

catégorielle montrent que la granularité de l'identification ne coïncide jamais avec celle de la discrimination (elle est plus grossière, [DH00]) : l'être humain est capable de discriminer au-delà de ce qu'il peut identifier, autrement dit, il est possible de discriminer à l'intérieur de catégories et pas simplement entre catégories<sup>27</sup> ;

- en situation de communication écologique<sup>28</sup>, ces seuils seraient probablement insuffisants pour permettre de détecter une différence de timbre.

Ces estimations nous seront néanmoins utiles lorsqu'il s'agira de comparer deux voyelles ayant un timbre proche. La variation induite par le contexte informationnel sera traitée plus loin en référence directe aux mots-test à structure /hVd/ que nous utilisons dans notre étude (Section 2.3.2). Mais nous pouvons toutefois affirmer pour l'instant que, en l'absence de contexte informationnel, le contraste entre les voyelles des mots-test ne repose que sur la qualité phonétique des timbres ; il est donc hautement probable que si deux voyelles ont des valeurs formantiques inférieures ou à peine supérieures aux seuils que nous venons de présenter, elles font probablement partie du même phonème.

Lorsqu'il s'agit de représenter graphiquement des voyelles à partir de mesures acoustiques, il est légitime de se demander si l'on souhaite ou non que la représentation fasse sens sur le plan de la perception. Nous n'entrerons pas dans les détails de la genèse des différentes échelles utilisées de nos jours car certains manuels d'introduction à la phonétique acoustique traitent le sujet (voir, par exemple, [Hay00], 140-145). Des études expérimentales ont abouti à deux échelles très employées en phonétique : les Bark (du nom du physicien allemand, Barkhausen) et les ERB (Equivalent Rectangular Bandwidth). L'Équation 2.1 permet la transformation des Hz en Bark que nous avons utilisée dans cette étude (adaptée de [Hay00], voir [Tra90] pour les détails). Cette Équation est valable pour  $x > 200$ .

---

<sup>27</sup>Pour être plus précis : Liberman et collègues ([LCSSK67]) soutiennent une version forte de la perception catégorielle pour les consonnes occlusives (la discrimination est à peine plus fine que l'identification) et une version plus faible – qui tend vers une perception continue – pour les voyelles (la variation intracatégorielle est aisément perçue).

<sup>28</sup>Par opposition aux conditions de laboratoire.

$$y = (26,81 \times x)/(1960 + x) - 0,53 \quad (2.1)$$

Intuitivement, si la taille des espaces vocaliques sur le plan F1/F2 varie d'un locuteur à l'autre en raison – entre autres – de différences d'ordre physiologique, on peut envisager qu'en procédant à une compression ou une expansion linéaire, les espaces individuels où seuls les aspects physiologiques sont facteurs de variation finissent par se ressembler. On sait d'ailleurs, d'après l'étude de Ladefoged et Broadbent ([Lad67]) que les valeurs acoustiques d'un son sont vraisemblablement jugées en référence à des caractéristiques acoustiques propres à chaque locuteur afin de déterminer l'appartenance de ce son à un phonème. En faisant varier les formants d'une phrase porteuse dans laquelle apparaît un monosyllabe test dont la voyelle reste constante, cette dernière est catégorisée différemment par les auditeurs. Par exemple, le stimulus synthétique <bit> a 375 et 1700 Hz pour F1 et F2, respectivement. Lorsqu'il est présenté après la phrase introductive « Please say what this word is » dans laquelle le premier formant varie de 275 à 500 Hz, environ 88 % de leurs 60 sujets identifient <bit>. Mais lorsqu'il est introduit par la même phrase avec un F1 variant de 200 à 380 Hz, 90 % des auditeurs entendent <bet>.

La question de la normalisation du locuteur a fait (et continue de faire) l'objet d'un grand nombre d'études (e.g. [Lob71], [Mil89], [RP94], [HR04], [ASvH04], etc.), et la plupart des travaux descriptifs basés sur des mesures acoustiques abordent des aspects de normalisation (e.g., [PB52], [Wel62], [Lad67], [Bla83], [HGCW95], [HM05], etc.).

Adank et collègues ([ASvH04]) évaluent 11 méthodes de normalisation en comparant leurs effets sur les 9 monophthongues du néerlandais produites par 160 locuteurs répartis en groupes sociolinguistiques. À partir de pourcentages de classification issus d'une série d'analyses discriminantes dont le but est de tester dans quelle mesure les procédures de normalisation préservent les contrastes phonémiques, réduisent la variation d'origine physiologique et maintiennent la variation sociolinguistique, les auteurs concluent que la méthode dite « de Lobanov » (centrer-réduire les fréquences des formants, indépendamment pour chaque formant), en référence à [Lob71], est la plus appropriée. C'est donc

cette méthode que nous utiliserons dans la suite de notre travail. <sup>29</sup>

### 2.3.2 Caractéristiques phonétiques des mots-test et information

La validité écologique des listes de mots du type *heed*, *hid*, *head* etc. est certes contestable dans une étude de phonétique, mais ces derniers offrent néanmoins la possibilité de comparer des voyelles en écartant les facteurs de variation importants que constituent le contexte consonantique et, plus généralement, le contexte informationnel (concept détaillé *infra*). Les variations de style ne sont pas abordées car elles ne sont pas spécifiques aux mots à structure /hVd/. En revanche, il est certain que le degré de formalité avec lequel les sujets ont abordé les enregistrements varie entre les individus, et il est, par conséquent, justifié de penser que ceci est un facteur supplémentaire de variation phonétique inter-individuelle. Une bonne compréhension de l'influence des contextes consonantique et informationnel permettra de mieux appréhender en quoi les formants mesurés en contexte /hVd/ peuvent se distinguer de ceux relevés en parole spontanée.

Pendant l'émission du [h], la langue n'étant pas requise comme articulateur dans la cavité supra-glottique, celle-ci peut donc anticiper la position de la voyelle subséquente. Par conséquent, l'effet de co-articulation de [h] sur la voyelle est quasi-nul. Le /d/ de la plupart des dialectes que nous étudions est une occlusive dont le lieu d'articulation correspond à la zone alvéolaire. Une étude de Delattre et collègues ([DLC55]) montre, à partir de stimuli synthétiques de type CV, que [d] possède un locus qui se situe aux alentours de 1800 Hz ; en d'autres termes, le deuxième formant de la voyelle suivant le [d]

---

<sup>29</sup>Pour rappel, l'écart-type d'une variable aléatoire discrète est une mesure de sa variation donnée dans l'Équation 2.2, où  $\bar{x}$  est la moyenne. Le centrage d'une variable consiste à soustraire  $\bar{x}$  et la réduction, à diviser le résultat de cette soustraction par  $s$  pour chaque valeur prise par cette variable (Équation 2.3). Les légendes des axes des figures de la Partie II désignent le centrage et la réduction d'une variable par le terme « z-score ».

$$s = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \quad (2.2)$$

$$z_i = \frac{(x_i - \bar{x})}{s} \quad (2.3)$$

tend vers une fréquence de 1800 Hz au fur et à mesure que l'on se rapproche de la barre d'explosion de la consonne. Une étude de Sussman et collègues ([SBDM97]) étudie l'effet du type de syllabe (CV, VC, CVC ou VCV) sur la valeur du locus. Par le biais d'une régression linéaire, les auteurs obtiennent une ordonnée à l'origine moyenne de 1232 Hz pour le cas qui nous occupe, i.e. le [d] en position de coda. Le type de syllabe semble ne pas avoir d'effet sur le locus de [d], contrairement à ce que les auteurs remarquent pour [b] et [g].

Le contexte informationnel regroupe, entre autres, les effets potentiels de la probabilité d'occurrence d'un mot<sup>30</sup>, de sa probabilité d'occurrence étant donnés les  $n$  mots qui le précèdent ou qui le suivent, ou encore étant donné le contexte sémantique ou situationnel et la position dans la phrase, sur la forme phonétique de ce mot. On sait par exemple (voir l'étude de Fidelholz rapportée dans [Jur03], 45) que la première voyelle des mots fréquents <forget> ou <forgive> est un meilleur candidat à la réduction vocalique que celle des items rares que sont <forgo> et <forfend><sup>31</sup>. La réduction vocalique est d'ailleurs souvent la norme pour les mots grammaticaux, qui ont la spécificité d'être particulièrement fréquents ([Byb01, BJFL<sup>+</sup>03, JBFL<sup>+</sup>98]). Lorsque van Bergem ([vB93]) montre l'effet de l'accent de phrase, l'accent de mot et la classe (lexicale *vs* grammaticale) sur la réduction vocalique phonétique en néerlandais, on peut considérer qu'il n'étudie en réalité qu'un seul facteur sous-jacent : l'information.

Les études qui ont examiné l'effet de la probabilité d'occurrence de plusieurs mots ([FLM99, BJFL<sup>+</sup>03]), qu'il s'agisse de  $P(m_i|m_{i-1})$ , la probabilité du mot  $m_i$  sachant que

---

<sup>30</sup>Les probabilités sont généralement déduites de fréquences d'occurrence des mots (nous utilisons ce dernier terme pour simplifier le propos) dans des corpus. L'estimation de cette probabilité ne correspond pas au rapport de l'effectif d'un mot donné sur le nombre total de mots d'un corpus ; en effet, ceci conduit à attribuer à tort une probabilité nulle aux mots absents du corpus et donc, *a contrario*, à surestimer les probabilités des items présents dans le corpus. Ce biais peut être atténué par une estimation de Good-Turing ([GS95], voir aussi [Jur03], 206 *sqq.*).

<sup>31</sup>Ces résultats ne disent toutefois pas si le processus est productif ou bien si la répartition entre voyelles réduites et non réduites a été fixée une fois pour toutes en diachronie. Il est d'ailleurs vraisemblable – et ceci répond peut-être en partie à la question – que ces quatre items ne suivent pas exactement le même schéma en anglais britannique et américain. En effet, les dictionnaires de prononciations ([Wel90] et [Jon03]) notent une alternative entre voyelle pleine et voyelle réduite en anglais américain pour la première syllabe de <forget> et <forgive>, mais une seule valeur phonologique (le schwa) pour l'anglais britannique.

le mot  $m_{i-1}$  a été prononcé, ou  $P(m_{i-1}m_i)$ , la probabilité d'occurrence des mots  $m_i$  et  $m_{i-1}$ , montrent un effet de cette probabilité sur la substance phonétique : les phénomènes de réduction (chute d'une consonne en coda, réduction vocalique, durée plus brève d'un mot) sont associés à des items dont la probabilité d'occurrence (ou de co-occurrence) est élevée. Par exemple, la probabilité conditionnelle plus élevée des séquences <did you>, <told you> et <would you> donne lieu à un nombre de palatalisations plus important des coronales que les paires <at you>, <but you> et <good you>, pour lesquelles  $P(m_i|m_{i-1})$  est moins élevé ([Jur03], 52). D'autres auteurs cités dans [Jur03], 53, ont démontré une corrélation entre la probabilité conditionnelle de bigrammes et la localisation de l'accent de phrase : plus la probabilité du bigramme est faible, plus la probabilité qu'il reçoive l'accent de phrase est importante.

En résumé, l'entropie linguistique, i.e. le degré d'incertitude associé à la réalisation d'un mot dans un certain contexte, affecte sa forme phonétique. Plus précisément, il apparaît qu'à une entropie faible correspond une réalisation phonétique plus déviante d'une réalisation prototypique (telle qu'elle est suggérée par la forme phonologique d'un mot), et inversement dans le cas d'une entropie élevée. On peut donc avancer que, lorsque l'occurrence d'un mot est totalement déterminée par le contexte, ou qu'elle est bien plus probable que celles d'autres mots dans ce même contexte (dans les deux cas, l'entropie est faible), sa forme phonétique a un degré de liberté accru par rapport à une forme canonique. La théorie H&H de production de la parole et le Neighborhood Activation Model (NAM) constituent un cadre particulièrement adapté à la description et la compréhension de ces phénomènes.

La théorie H&H, proposée par Lindblom ([Lin90]), met l'accent sur la relation entre la production du locuteur et ce que ce dernier anticipe des attentes de l'auditeur pour que l'accès lexical soit optimal. Optimal... et non maximal<sup>32</sup>, car la théorie H&H prévoit une

---

<sup>32</sup>D'ailleurs, même dans des conditions d'entropie élevée, incitant le locuteur à hyper-articuler, sa production n'ira jamais jusqu'à satisfaire pleinement les attentes de l'auditeur, comme le prouve l'effet « hyperspace » ([JFW93, Joh00]) : lorsque des sujets sélectionnent des voyelles de synthèse prototypiques de leur système, l'espace acoustique obtenu est plus grand que celui qui est généré par des mesures acoustiques de voyelles prototypiques produites par les mêmes sujets. Si Whalen et collègues ([WMP<sup>+</sup>04])

interaction et la recherche constante d'un équilibre entre :

- le contexte informationnel, ou, dans la théorie H&H, les « signal-complementary processes » ;
- les contraintes « output-oriented » ; il s'agit d'optimiser l'information acoustique pour faciliter l'accès lexical. C'est l'« hyperspeech », auquel le locuteur a recours lorsque l'entropie est élevée<sup>33</sup> ;
- le contrôle « system-oriented », qu'on peut assimiler à la notion d'économie de l'articulation : « [u]nconstrained, a motor system tends to default to a low-cost form of behavior. » ([Lin90], 413).

La forme acoustique, qui dépend de l'équilibre entre ces trois paramètres, est donc censée constituer un contraste optimal. Reste à définir comment ce contraste est implémenté dans le domaine de la production, en particulier, sur le plan acoustique. Une très brève description du Neighborhood Activation Model (NAM) est instructive, et adaptée à notre propos en vertu du fait que ce dernier rend compte des effets de fréquences, et de l'entropie, qu'il est cité par Lindblom ([Lin90], 409-410), et que ses prédictions ont été utilisées dans une étude acoustique d'un grand intérêt pour la suite de notre propos.

Le modèle NAM ([GPL96], 311-313) prévoit que les mots fréquents ont une plus forte probabilité d'être reconnus, et de l'être rapidement, mais cette probabilité est pondérée à la baisse par :

- la densité du voisinage, i.e. le nombre de mots possédant des caractéristiques phonético-acoustiques proches du stimulus et leur degré de confusion possible avec le stimulus ;
- la fréquence du voisinage : la fréquence des candidats du voisinage par rapport à la fréquence du stimulus.

Si Goldinger et collègues ([GPL96], 311-313) rapportent que ces deux aspects affectent la reconnaissance des mots, il serait donc informatif, si l'on se replace dans le cadre de la théorie H&H, de connaître ce que le locuteur anticipe des difficultés de l'auditeur

---

répliquent en partie – mais de façon flagrante – cet effet, ils émettent quelques réserves quant à l'utilité du concept d'« hyperspace » pour expliquer l'apparent décalage entre production et perception.

<sup>33</sup>Nous laissons de côté l'adaptation du locuteur au bruit environnant, parfois appelée « effet Lombard » car, bien que cet aspect fasse partie du contexte informationnel, nous postulons que les effets de ce type de bruit sont constants dans nos données.

à reconnaître les mots en contexte /hVd/. Malheureusement, tel qu'il est décrit dans [GPL96], le modèle NAM ne règle pas la question de la distance entre le stimulus et ses voisins. S'agit-il d'une simple distance de type Levenshtein ([DHS01], 418-420), où les unités primitives sont les phonèmes, ce qui aboutirait, par exemple, à considérer que la distance entre /hi:d/ et /hɪd/ est la même que celle qui sépare /hi:d/ et /hæd/? On peut encore imaginer une distance du même type, mais cette fois entre des traits phonologiques, ou peut-être une distance phonétique euclidienne dans l'espace des trois premiers formants et de la durée. Quelle que soit la forme de cette distance, on doit pouvoir considérer qu'elle contraint le degré de liberté acoustique des listes de mots en /hVd/. En effet, si l'on pose que la liste de 19 mots à structure /hVd/ constitue un lexique de taille finie où les items sont équiprobables<sup>34</sup>, on est dans la situation où la densité du voisinage et la fréquence du voisinage sont extrêmement élevés. Par conséquent, du point de vue de l'auditeur, l'accès au lexique se trouve compliqué, et du point de vue du locuteur, celui-ci doit tout mettre en œuvre pour maximiser les contrastes<sup>35</sup>. Il faut donc s'attendre à ce que les distances acoustiques entre deux mots-test soient magnifiées par rapport à de la parole spontanée. Notons en outre que le recours au décodage en phonèmes dans la reconnaissance lexicale est vraisemblablement cantonné à certains cas bien précis. D'après Sendlmeier ([Sen95], 140) :

a listener can switch to single sounds or even distinctive features when discriminating, for example, minimal pairs or difficult words such as proper names, words of a foreign language or pseudowords, and he/she can then switch back again to the syllable or even word level when progressing in the recognition process.

Le recours au phonème comme unité primitive pour l'accès au lexique semble donc s'appliquer aux cas où l'entropie est élevée (paires minimales hors contexte, noms propres, pseudo-mots, etc.), ce qui est le cas dans la liste des mots /hVd/ de notre corpus. À redondance linguistique élevée, la reconnaissance lexicale semble ne plus impliquer le phonème ; produire un prototype dans ce cas-là devient non seulement inutile, mais également dis-

---

<sup>34</sup>On peut en effet concevoir que les locuteurs ont rapidement réalisé que leur tâche consistait à prononcer un faible nombre de mots ayant un contexte consonantique constant. Si l'on écarte le cas de *hured*, où la correspondance entre forme phonologique et graphique est loin d'avoir fait l'objet d'un consensus, nous doutons que les probabilités d'occurrence différentes de chacun des mots-test dans la parole naturelle affecte leur forme phonétique dans cette tâche artificielle.

<sup>35</sup>Ceci est probablement contrebalancé par le fait qu'une liste de mots est un cas limite, peu écologique, d'acte de communication.

pendieux en termes de coût articulaire, comme le prévoit la théorie H&H. Il s'ensuit que la distance que le locuteur cherche à maximiser dans le cas des listes de mots en /hVd/ concerne très probablement le phonème (voire une entité de taille inférieure)<sup>36</sup>.

Aylett ([Ayl00]) reformule la théorie H&H dans un cadre plus proche de la théorie de l'information. Pour ce faire, il distingue trois types de redondances :

- la redondance linguistique, qui traduit la proportion des entités (mots, syllabes, phonèmes, etc.) dont la présence est statistiquement déterminée dans un message. C'est dans cette acception que le terme est utilisé par Shannon et Weaver ([SW63], 43-44) ; par exemple, <to> dans <I'm going to leave you> est très redondant ;
- la redondance acoustique : à contexte informationnel égal, un mot est plus facilement identifiable si sa redondance acoustique est élevée. Cette redondance se traduit par une articulation plus soignée, très vraisemblablement synonyme du « hyperspeech » de Lindblom ([Lin90]) ;
- la redondance du signal : il s'agit de la combinaison des deux types de redondance que nous venons de mentionner. Aylett formule l'hypothèse suivante : la redondance du signal tend à varier le moins possible au cours du temps. Autrement dit, il prévoit que la faible redondance d'une entité linguistique sera compensée par une redondance acoustique accrue, et, inversement, la redondance linguistique élevée d'une entité ne nécessitera qu'une faible redondance acoustique pour remplir le critère de variation minimale de la redondance du signal au cours du temps.

Le lien avec la théorie H&H est évident, surtout si l'on admet que la redondance acoustique est proportionnelle à l'effort d'articulation. Jusqu'ici, les études empiriques auxquelles nous avons fait allusion confirment ces deux formulations de la théorie H&H. Mais elles n'apportent rien à la compréhension de la manière dont les formants des voyelles de nos mots-test peuvent être affectés. En effet, les résultats décrits jusqu'ici sont binaires :

---

<sup>36</sup>La question de l'unité primitive de perception de la parole n'est toujours pas résolue ([GA03]) ; par exemple [MWW94] soutient l'interprétation directe des traits phonétiques, sans passer par le niveau phonémique. Dans une autre perspective, Durand et Laks ([DL02], 30) désignent la syllabe comme unité fondamentale « which provides the link between a physically driven system and higher-level units ». Ce que nous venons de décrire est donc, certes fondé empiriquement et intuitivement, mais nous ne prétendons pas qu'il s'agit d'une réponse définitive.

soit la voyelle est pleine, soit elle est réduite ; soit la consonne est produite, soit elle est élidée. Or les mots-test à structure /hVd/ sont des formes isolées accentuées ; leur voyelle ne peut donc pas être candidate à la réduction phonologique.

Wright ([Wri03]) analyse dans quelle mesure la densité du voisinage et la fréquence relative d'un stimulus affectent les valeurs des formants. Il utilise deux types de stimuli :

- les mots « faciles », qui ont un voisinage peu dense et une fréquence élevée par rapport à leurs voisins ;
- les mots « difficiles », avec les caractéristiques opposées : un voisinage dense et une fréquence relative faible.

F1 et F2 sont ensuite mesurés au point temporel où ils atteignent une distance maximale par rapport au centre de l'espace vocalique, puis transformés en Bark. Conformément aux études précédentes (e.g. [Lin83]), la taille de l'espace vocalique délimité par les timbres moyens de la condition « mots difficiles » est supérieure à celle de l'espace des « mots faciles ». Mais le résultat le plus intéressant est le suivant : la dispersion autour de la moyenne des valeurs de chaque timbre est plus importante (pour la majorité des voyelles) dans la condition « difficile ». Les différences de dispersion les plus marquées entre les deux conditions concernent les voyelles situées aux extrémités de l'espace vocalique ; et ce sont celles-là même qui exhibent la différence moyenne la plus saillante entre les deux conditions. Autrement dit, l'augmentation moyenne de distance entre les timbres est accompagnée d'une augmentation de leur variance individuelle ; mais l'étude ne précise pas explicitement si les contrastes sont améliorés ou non, i.e. si les frontières entre les nuages de dispersion de timbres dans F1/F2 sont plus ou moins nettes.

Aylett et Turk ([AT06]) ont des résultats comparables, qui confirment dans l'ensemble l'hypothèse de la centralisation dans F1/F2 des voyelles associées à une redondance élevée. Ils ajoutent que la présence de proéminences prosodiques est associée à des syllabes de faible redondance linguistique ; la prosodie permet donc elle aussi de moduler la redondance acoustique en fonction de la redondance linguistique afin d'assurer la stabilité de la redondance du signal dans le temps.

À ce stade de la discussion, il apparaît que la variation acoustique est dans une large mesure déterminée par des différences physiologiques et par le contexte informationnel. Ces dernières seront en partie écartées par les méthodes dites « de normalisation » employées dans la Section 6.2. Le contexte informationnel est, quant à lui, de fait neutralisé par l'utilisation des mots à structure /hVd/, ce qui constitue un atout majeur pour une étude des propriétés acoustiques des voyelles. Nous pouvons donc affirmer que les mots à structure /hVd/ constituent un outil d'étude des dialectes particulièrement adapté (ceci sera d'ailleurs mis en évidence par les scores de classification obtenus dans la Section 7.2) car ils offrent un contrôle des facteurs qui est optimal pour une étude observationnelle. Ils comportent néanmoins certaines faiblesses qui ne sont pas dues au fait qu'il s'agisse de formes isolées, mais plutôt à l'exhaustivité de la liste de ces mots-test, au degré de consensus avec lequel les locuteurs vont attribuer une forme phonologique à certains mots rares ou pseudo-mots (nous mentionnons le cas de *hured* dans la Section 6.3.1.3) et à l'impossibilité de garantir *a priori* l'appartenance d'un mot-test à un ensemble lexical donné. En ce qui concerne l'exhaustivité de la liste, les 19 mots disponibles ne couvrent pas l'ensemble des contrastes possibles dans les 13 dialectes étudiés. Par exemple, le phénomène de convergence COT-CAUGHT<sup>37</sup>, caractéristique, notamment, de certaines variétés de l'anglais d'Écosse ne peut pas être mis en évidence puisque aucun des mots de la liste de ne peut être assimilé à l'ensemble CAUGHT. En effet, le seul mot-test présentant *a priori* une voyelle postérieure mi-fermée est *hoard* ; or s'il peut, faute de mieux, faire l'affaire pour tester une éventuelle convergence COT-CAUGHT (non attestée) en anglais standard par exemple, ce n'est pas le cas en anglais d'Écosse car la rhoticité fait de *hoard* un mauvais candidat pour l'ensemble CAUGHT. Les mots <wait> et <weight> ne riment pas dans certains dialectes du nord de l'Angleterre ([ei] et [ɛi], respectivement) ; or il n'est pas possible de révéler ce contraste avec le mot-test *hade*, représentant de la supra-catégorie FACE. Il arrive également que la variation allophonique ait un intérêt dans la description des dialectes. En effet, nous verrons plus loin (Section 3.2.1) que l'ensemble PRICE

---

<sup>37</sup>On rencontre souvent cette dénomination bien que les ensembles lexicaux standards soient LOT et THOUGHT.

peut-être subdivisé en deux sous-ensembles à Hull et Newcastle (la qualité de la voyelle ou son degré de diphtongaison étant affecté par le voisement de la consonne qui suit), et ici encore, le mot-test *hide* ne permet pas de révéler cette variation supplémentaire (par rapport à l'anglais standard).

Il convient maintenant d'examiner plus précisément la manière dont l'espace multidimensionnel des paramètres physiques est projeté dans un espace de catégories phonémiques.

### **2.3.3 Théorie du prototype et théorie des ensembles flous**

Dans la pensée structuraliste ([Sau95, Jak63]), le système phonologique d'une langue contient des entités discrètes formant des catégories hermétiques les unes aux autres. Pour accéder au statut d'entité pertinente, de phonème, un son doit constituer un contraste fonctionnel qu'il est possible de révéler par le biais du test des paires minimales. Les catégories regroupent les formes abstraites de sons définis par des traits distinctifs. Pour appartenir à une catégorie, une entité doit en posséder tous les traits distinctifs et les posséder pleinement. Il n'est pas ici question de rejeter ce point de vue, mais il nous incombe de trouver une solution alternative nous permettant de résoudre l'apparent paradoxe selon lequel un seul et même phonème a pour équivalents tangibles des événements acoustiques très variables – mais, comme nous venons de le voir, cette variabilité est en grande partie prévisible. La théorie du prototype constitue un cadre tout à fait adéquat dont nous introduisons quelques rudiments avant d'aborder cette notion par le biais du cadre formel de la théorie des ensembles flous.

Un prototype désigne souvent le membre d'une catégorie qui est (quasi) universellement considéré comme un bon représentant de celle-ci. Par exemple, les études sur la catégorisation des couleurs par Berlin et Kay, Eleanor Rosch et d'autres (rapportées dans [Tay95], 1-20) font apparaître que certaines couleurs – 11 pour Berlin et Kay – forment un inventaire que l'on retrouve dans presque toutes les langues étudiées. Ces couleurs, qui ont donc un statut particulier, sont qualifiées de « focales » ; elles exhibent, en outre, les

caractéristiques suivantes (d'après Rosch) :

- elles sont stables quelle que soit la langue parlée : lorsqu'on demande à des sujets de cultures diverses de choisir un bon exemple de rouge, ils ont tendance à retenir la même nuance ; il existerait donc une nuance de rouge qui puisse probablement être universellement considérée comme le centre de la catégorie « rouge » ;
- une expérience impliquant des locuteurs natifs de 23 langues différentes, à qui on a demandé de nommer les couleurs présentées sur des échantillons, répondent plus rapidement lorsqu'il s'agit de couleurs focales ;
- dans une autre expérience, une couleur est présentée à des sujets pendant 5 secondes ; puis, après 30 secondes, leur tâche consiste à reconnaître la couleur qu'ils ont vu parmi d'autres ; les couleurs focales sont reconnues plus facilement. Un aspect particulièrement intéressant de cette expérience réside dans le fait que les sujets Dani (tribu papoue de Nouvelle-Guinée), ne disposant que de deux termes pour tout le spectre des couleurs, montrent également de meilleures performances pour les couleurs focales.

Si ces catégories de couleurs « focales » semblent faire l'objet d'un consensus entre les cultures, il est fort probable que leur existence soit une conséquence de caractéristiques inhérentes des systèmes perceptif et cognitif de l'être humain. Comme le formule Taylor ([Tay95], 52) : « colour categories pre-exist their linguistic encoding »<sup>38</sup>. D'ailleurs, ce caractère « focal » ne se limite pas aux couleurs. Il inclut, entre autres, les formes géométriques, avec probablement une interaction entre facteurs purement biologiques et facteurs culturels<sup>39</sup>. Il semble en outre que certaines voyelles obéissent à ce principe. En effet, la description de Stevens ([Ste89]) des relations entre les domaines articulaire, acoustique et auditif peut être interprétée comme une description des raisons naturelles permettant l'émergence de prototypes de type focal, i.e. universellement contraints. L'auteur démontre que ces relations ne sont pas monotones ; c'est ce qui est illustré dans la

---

<sup>38</sup>Voir aussi Lakoff ([Lak87], 24-30).

<sup>39</sup>Voir, par exemple, l'étude de Raghurir et Greenleaf ([RG06]) sur les dimensions préférées des rectangles, et la référence au fameux rectangle d'or.

Figure 2.1. Dans la zone 1, le mouvement articuloire n'a que peu d'impact sur le signal acoustique, puis, dans la zone 2, le mouvement articuloire (pourtant moindre que celui de la zone 1) engendre une modification importante de la valeur que prend la variable acoustique ; enfin, la zone 3 illustre le même schéma que la zone 1. Les zones où le paramètre acoustique reste relativement insensible aux perturbations affectant le paramètre articuloire ont, d'après Stevens, des conséquences sur la formation de l'inventaire phonémique des langues. Donc, de même que les cônes de la rétine humaine sont naturellement plus sensibles à certaines longueurs d'ondes (phénomène avancé pour l'existence des couleurs focales, [Tay95], 13-14), ces zones de stabilité acoustique constituent de véritables « niches » propices à l'apparition de prototypes de voyelles.

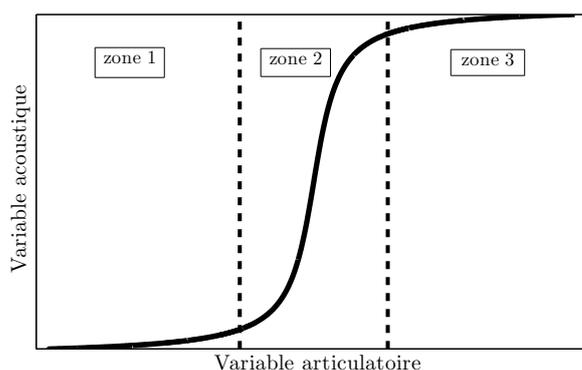


FIG. 2.1 – Illustration de la théorie quantique (adapté de [Ste89], 4).

Les prototypes auxquels nous faisons référence dans cette étude ne sont pas de type focal (tel que nous venons de définir ce terme), en cela qu'ils ne font pas référence à des représentants universellement jugés comme bons d'une catégorie. Ils sont dépendants du système vocalique de la langue, du dialecte, voire du locuteur. En d'autres termes, si la réalisation du /u/ de l'anglais britannique standard ([u]) n'est pas universellement considérée comme un exemple typique de [u], c'est néanmoins le prototype du /u/ de l'anglais britannique standard. La notion de prototype que nous utilisons ici est spécifique à une communauté linguistique précise, et est donc différente du prototype « focal » universel et biologiquement contraint. Puisque nos travaux s'appuient sur des mesures de grandeurs

physiques, le prototype dans notre étude n'est pas une représentation abstraite du centre d'une catégorie, mais bel et bien un « exemplaire »<sup>40</sup>. En d'autres termes, dans notre étude, le prototype d'une catégorie phonémique est un point dans un espace multidimensionnel dont les dimensions sont des variables corrélées aux grandeurs physiques mesurables dans le signal. Pour prendre un exemple simplifié, si l'on considère que la valeur des deux premiers formants au milieu temporel de la voyelle décrit de façon exhaustive la manière dont les voyelles s'organisent en catégories, on pourrait considérer que la valeur médiane<sup>41</sup> de chaque phonème représenté dans le plan F1/F2<sup>42</sup>, est le prototype de cette catégorie. Autrement dit, le degré d'appartenance d'une voyelle à cette catégorie peut-être directement estimé par une mesure appropriée de la distance entre les coordonnées de cette voyelle et celles de la voyelle possédant les valeurs médianes dans le plan F1/F2. Cette dernière voyelle n'est donc pas une simple « instanciation » d'un prototype abstrait ; elle est le prototype.

Le prototype « appris » (par opposition à focal) d'un phonème a dans une langue donnée présente certaines caractéristiques intéressantes. Comme le résumait Iverson et Kuhl ([IK95], 553) :

Stimuli judged as exceptionally good instances of phonetic categories (prototypes) make neighboring tokens in the vowel space seem more similar, exhibiting a *perceptual magnet effect*.

Dans l'étude de Kuhl ([Kuh91]), l'auteur a synthétisé plusieurs exemples de /i/, et les a soumis à des sujets américains adultes afin qu'ils les jugent. Puis, le timbre perçu comme le meilleur /i/ ainsi que le plus mauvais ont été choisis, et leurs formants manipulés afin de créer des variantes de ces deux stimuli de référence. Une tâche de discrimination impliquant des sujets adultes et des nourrissons a démontré ce qui est depuis connu sous le nom de « perceptual magnet effect » : les variantes du prototype sont plus difficiles à discriminer que ne le sont les variantes du timbre non prototypique. D'autres études (citées dans [IK95], 553) ont ensuite démontré que :

---

<sup>40</sup>Traduction littérale de l'anglais <exemplar>.

<sup>41</sup>Il s'agit ici de la médiane comme estimateur de la tendance centrale des distributions de F1 et F2.

<sup>42</sup>Voir les illustrations de la Partie II : Figure 6.12 et suivantes du même type.

- il est attesté chez le nourrisson dès l'âge de six mois ;
- le prototype, dans ce cas-là est une entité apprise, propre à un système phonologique, comme le démontrent des études translinguistiques ;
- il implique une distorsion de l'espace acoustique qui conduit à une réduction de la distance perceptive autour d'un prototype.

Ceci a des conséquences directes sur la manière dont les catégories sont représentées, et, plus particulièrement, sur la variation acoustique tolérée à l'intérieur d'une catégorie phonémique. En effet, ce phénomène peut être interprété comme un degré de liberté acoustique supplémentaire car la distance perceptive d'un stimulus au prototype est virtuellement nulle quand sa distance objective est très faible. Dans la discussion de leurs résultats, Iverson et Kuhl ([IK95], 560) insistent sur l'apport de cet effet attracteur au domaine de la perception catégorielle : il avait déjà été démontré que la discrimination de voyelles aux frontières de catégories était plus aisée que celle impliquant des timbres à l'intérieur de catégories. L'effet attracteur du prototype démontre en outre qu'il existe à l'intérieur d'une seule et même catégorie un degré de conformité au prototype qui influence la distance perçue entre deux exemplaires de voyelles.

Broad ([Bro76]) aborde la question des catégories par le biais de la production et non pas de la perception. Son approche est intéressante car, bien que la théorie des ensembles flous et la théorie des probabilités offrent des perspectives différentes ([Zad95]), sa description du chevauchement des catégories par le biais de gaussiennes dont il ajuste de manière heuristique les paramètres pour les besoins de sa démonstration, n'est pas très éloignée, conceptuellement, de la philosophie des ensembles flous. À partir de la dispersion des trois premiers formants de voyelles répétées par un même locuteur, il partitionne l'espace vocalique en plusieurs sous-espaces incluant 1,28 écart-type au-dessus et au-dessous de la moyenne (i.e. ces sous-espaces incluent 80 % de la distribution théorique). En première approximation, cela aboutit à des valeurs, à peu près constantes pour toutes les voyelles, de 40, 140 et 160 Hz pour F1, F2 et F3 (ces valeurs concernent des voyelles isolées). Ces résultats donnent une idée de la taille des catégories phonémiques dans l'espace des trois

premiers formants.

La catégorisation des voyelles dans un contexte multi-dialectal a été étudiée par Evans et Iverson ([EI04]). Dans cette étude, les sujets jugent des voyelles dans des monosyllabes présentés à l'intérieur d'une phrase porteuse, tantôt prononcée avec un accent du nord, tantôt avec un accent du sud-est. Dans une première expérience, deux groupes de sujets, vivant tous à Londres, dont l'accent d'origine est soit du nord, soit du sud-est, ont pour tâche de déterminer la meilleure voyelle dans des mots synthétisés tels que <bud> ou <bath> (ces stimuli comportent les traits déterminant la partition nord/sud de l'Angleterre). Lorsque la phrase porteuse est produite dans l'accent d'origine des sujets, ceux-ci choisissent les stimuli dont les formants sont en adéquation avec la production attendue d'un locuteur de cet accent. Par exemple, pour <bud> présenté dans une phrase porteuse dans leur accent d'origine, les sujets originaires du sud choisissent [ʌ] et ceux du nord, [ʊ]. Lorsque la phrase n'est pas produite dans leur accent d'origine, les sujets des deux groupes ne choisissent aucune des ces deux variantes, mais plutôt une voyelle intermédiaire de type [ə]. Dans un second temps, un troisième groupe de sujets, composé d'adolescents entre 16 et 17 ans originaires de Ashby de la Zouch (nord linguistique de l'Angleterre), est soumis au même test. Leur meilleur exemplaire, quelle que soit la phrase porteuse, est celui qui correspond à la production attendue d'un locuteur septentrional. Ceci démontre que :

- des locuteurs vivant à Londres, et étant de ce fait en contact avec des accents du nord et du sud, sont capables d'adapter leur représentation prototypique d'un timbre vocalique. Cette représentation est toutefois en partie erronée (cas du [ə]) lorsque l'accent de la phrase porteuse est différent de leur accent d'origine.
- des locuteurs n'ayant été que très peu en contact avec un accent différent du leur ne procèdent pas à cette adaptation au dialecte.

Le premier point démontre que des auditeurs en contact prolongé avec d'autres accents ajustent leur représentation prototypique d'un timbre en fonction du dialecte du locuteur. De plus, cet ajustement ne semble pas être conforme à la manière dont l'apprenant assimile les phonèmes d'une langue étrangère à ceux de sa langue maternelle. En effet, les valeurs

intermédiaires entre [ʌ] et [ʊ] obtenues dans le cas où l'accent d'origine des sujets était différent de celui de la phrase porteuse tendent à prouver, d'après Evans et Iverson, que la constitution éventuelle d'un système phonologique dialectal supplémentaire passe par la formation de prototypes à partir d'épisodes stockés dans la mémoire à long terme, et non par l'assimilation des voyelles du nouveau système à celles de celui du dialecte natif. L'absence de correspondance exacte entre ce prototype nouvellement formé et les productions typiques du dialecte pourrait être imputable à la distorsion induite par l'effet attracteur des prototypes du dialecte d'origine.

Dans leur étude suivante, Evans et Iverson ([EI07]) poussent davantage leur analyse. Ils inspectent l'évolution des productions de locuteurs originaires du nord sur leurs deux premières années à l'université, période pendant laquelle, selon les auteurs, ils sont confrontés à l'anglais standard (du sud). L'hypothèse de départ est que le système vocalique des sujets va tendre vers celui de l'anglais standard. Dans un premier temps, les sujets sont enregistrés à 4 points temporels sur les deux ans, et leur production est jugée par 6 phonéticiens sur un continuum nord/sud. Les résultats font apparaître que, malgré des différences individuelles importantes (certains locuteurs avaient au départ un degré de proximité au pôle « nord » très différents), la tendance générale montre un déplacement du nord vers le sud. Les auteurs se penchent ensuite sur la reconnaissance de phrases dans le bruit : une série de phrases sont présentées aux sujets, tantôt lues avec un accent du nord, tantôt avec en anglais standard, avec différents rapports signal/bruit. Les résultats montrent que les sujets originaires du nord, mais dont le système vocalique atteint les meilleurs scores de proximité au sud dans l'expérience décrite plus haut, identifient mieux les phrases prononcées en anglais standard.

Compte tenu des éléments passés en revue dans cette section, la formalisation de la théorie du prototype par le biais de la théorie des ensembles flous semble justifiée. Nous développons dans ce qui suit un cadre conceptuel qui, sans apparaître explicitement dans la suite de notre travail (sauf Sections 7.6 et 7.7), a sous-tendu notre vision de l'interface phonétique/phonologie.

Nous nous basons notamment sur Negnevitsky ([Neg02], 87-128) et Jang et Gulley ([JG99]); la description originale de la théorie est exposée dans Zadeh ([Zad65]). Notre exemple, simplifié, est le suivant : les voyelles peuvent être antérieures, centrales ou postérieures ; comment caractériser l'appartenance d'une voyelle à l'une de ces trois catégories à partir des valeurs de F2 ? Soit  $X$ , l'univers du discours (i.e. ici, les fréquences de F2 centrées-réduites<sup>43</sup>), contenant quatre éléments (voyelles) :  $X = \{x_1, x_2, x_3, x_4\}$ . Le sous-ensemble  $A$  (les voyelles antérieures) de  $X$  contient deux éléments :  $A = \{x_1, x_2\}$ . En logique binaire classique, le sous-ensemble  $A$  peut-être décrit de la manière suivante :  $A = \{(1/x_1), (1/x_2), (0/x_3), (0/x_4)\}$ , où chaque élément  $x_i$  est caractérisé par sa fonction d'appartenance  $\mu_A(x_i)$  au sous-ensemble  $A$  sous la forme  $\{\mu_A(x_i)/x_i\}$ . Dans ce cas,  $\mu_A(x_i)$  ne peut prendre que deux valeurs :

1.  $\mu_A(x_i) = 1$ , si  $x_i \in A$  ;
2.  $\mu_A(x_i) = 0$ , si  $x_i \notin A$ .

Ceci conduit à la situation illustrée dans la Figure 2.2<sup>44</sup>, qui va à l'encontre de l'intuition et de ce que nous venons de voir concernant la catégorisation. On y constate que tout  $x$  ayant une fréquence  $Freq$  comprise entre environ 0,833<sup>45</sup> et 2,5 écart-types appartiennent pleinement et exclusivement au sous-ensemble  $A$ , et subitement, à partir de  $Freq < 0,833$ ,  $x$  ne fait plus du tout partie du sous-ensemble  $A$ .

La théorie des ensembles flous propose une *via media* évitant cette rupture abrupte, particulièrement aberrante lorsque l'univers du discours prend la forme d'une variable continue<sup>46</sup>. Dans ce contexte,  $\mu_A(x_i)$  peut prendre comme valeur tout nombre réel entre 0 et 1 compris. La Figure 2.3 représente une première tentative de modélisation des catégories antérieur ( $A$ ), central ( $C$ ) et postérieur ( $P$ ) par le biais d'ensembles flous. Les

---

<sup>43</sup>Pour l'exemple, les valeurs de F2 sont bornées entre -2,5 et 2,5 écart-types.

<sup>44</sup>Toutes les figures dans ce travail ont le point comme séparateur de décimales ; la virgule est utilisée partout ailleurs.

<sup>45</sup>Cette valeur délimite le premier tiers de l'univers du discours.

<sup>46</sup>Sauf, peut-être, dans les cas avérés de perception catégorielle au sens fort, comme, par exemple, la perception de [b], [d] et [g] à partir de modifications progressives des transitions de F2 ([LCSSK67], 442 *sqq.*), qui correspondent, sur le plan de la perception, à des sauts quantiques d'une catégorie à l'autre.

triangles<sup>47</sup> représentent les fonctions d'appartenance d'une voyelle à l'une des trois catégories. Par exemple, on sait que  $x_1 \in A$ ; en imaginant que  $x_1 = 1,3$  dans l'univers du discours, la ligne discontinue schématise le degré d'appartenance de  $x_1$  au sous-ensemble  $A$ , et la ligne en pointillés, le degré d'appartenance de  $x_1$  à  $C$ . On peut donc noter que  $\mu_A(x_1) = 0,4146$  et  $\mu_C(x_1) = 0,1333$ . La voyelle correspondant à  $x_1$  est donc partiellement antérieure, et centrale dans une moindre mesure.

Un avantage supplémentaire de la théorie des ensembles flous réside dans le fait qu'il est possible de créer de nouveaux sous-ensembles, comme l'être humain le fait tous les jours, par le biais de variables linguistiques, ce que les anglophones nomment « hedges ». Par exemple, les adverbes d'intensité permettent une telle modification : les hommes peuvent être grands, très grands, extrêmement grands, plus ou moins grands, etc. De même, il est possible de considérer qu'une voyelle peut être perçue comme très centrale, plus ou moins centrale, etc. La théorie des ensembles flous met à disposition des expressions mathématiques pour tenter de capturer la manière dont ces sous-ensembles formés à partir de « hedges » peuvent être dérivés des fonctions caractérisant des sous-ensembles de départ. Dans la Figure 2.4, nous donnons l'exemple du sous-ensemble des voyelles très centrales et de celui des voyelles plus ou moins centrales, obtenus, en suivant [Neg02], 97 :

$$\mu_C^{tres}(x) = [\mu_C(x)]^2 \quad (2.4)$$

$$\mu_C^{+/-} = \sqrt{[\mu_C(x)]} \quad (2.5)$$

Si l'on pose que le prototype  $x_{proto}$  d'une catégorie (ou sous-ensemble)  $A$  est caractérisé par le degré d'appartenance  $\mu_A(x_{proto}) = 1$ , il s'ensuit que, en utilisant une fonction triangulaire,  $x_{proto}$  ne peut prendre qu'une seule valeur dans l'univers du discours. Autre-

---

<sup>47</sup>Ces fonctions sont conventionnellement qualifiées de triangulaires, mais il ne s'agit là que d'une image facilitant la visualisation des sous-ensembles. Il est par exemple évident que, si le triangle qui matérialise le sous-ensemble  $C$  était un véritable triangle (i.e. si tous les sommets étaient reliés), la fonction de  $x_1$  prendrait deux valeurs : 0 et 0,1333 (le point  $x_1$  est illustré dans la Figure 2.3).

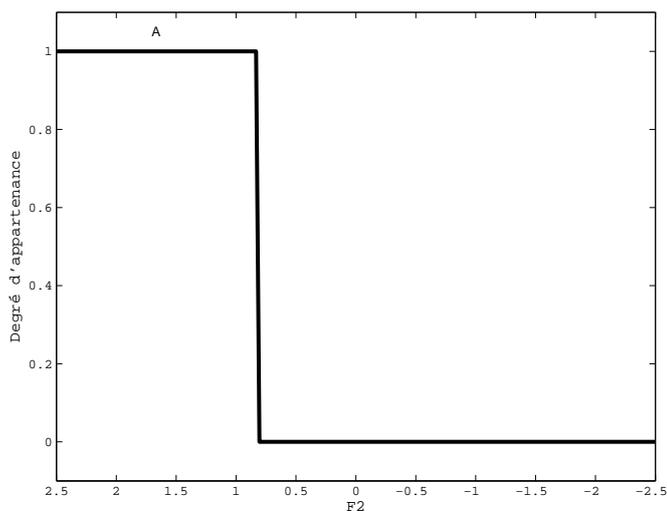


FIG. 2.2 – Illustration des limites de la logique binaire pour la catégorisation.

ment dit, cela impliquerait que la voyelle prototypique d'une des trois catégories de notre espace unidimensionnel est exclusivement représentée par une seule valeur de F2 et que le degré de ressemblance au prototype des autres  $x_n$  voyelles décroît de façon linéaire au fur et à mesure que  $|Freq(x_i) - Freq(x_{proto})|$  croît. Ceci soulève deux problèmes : d'une part, étant donné le degré de liberté des valeurs formantiques, en partie imputable aux limites de résolution de l'appareil auditif (comme nous l'avons noté plus haut) on peut considérer intuitivement que, lorsque  $|Freq(x_i) - Freq(x_{i+1})|$  est très faible,  $\mu_A(x_i) \approx \mu_A(x_{i+1})$ . D'autre part, étant donné l'effet attracteur des prototypes que nous avons mentionné plus haut, on peut concevoir qu'en deçà d'une certaine valeur de  $|Freq(x_i) - Freq(x_{proto})|$ , le degré d'appartenance à la catégorie décroît très lentement pour ensuite, passée cette valeur, diminuer de façon plus abrupte. Cette notion est illustrée dans la Figure 2.5 ; les fonctions triangulaires ont été remplacées par des sigmoïdes<sup>48</sup>.

La théorie des ensembles flous permet de modéliser la correspondance entre les paramètres acoustiques et les catégories phonémiques prototypiques. Un intérêt supplémentaire réside dans le fait que le degré d'appartenance aux catégories reflète à bien des égards le jugement humain. Nous n'utilisons pas ce type de représentation dans ce qui suit, mais il

<sup>48</sup>La courbe représentant le sous-ensemble  $C$  est obtenue grâce à la fonction `psigmf` de la Fuzzy Logic Toolbox du logiciel Matlab ; celles des sous-ensembles  $A$  et  $P$  utilisent la fonction `dsigmf` (voir le guide de l'utilisateur pour les détails).

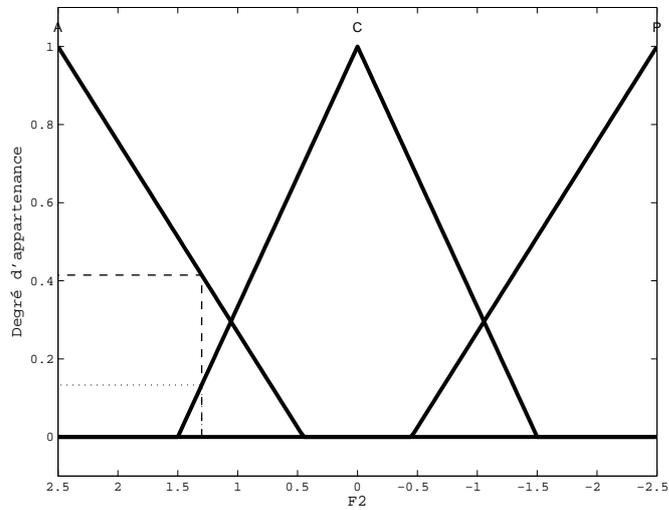


FIG. 2.3 – Illustration du degré d'appartenance d'un élément à un sous-ensemble dans le cadre de la théorie des ensembles flous.

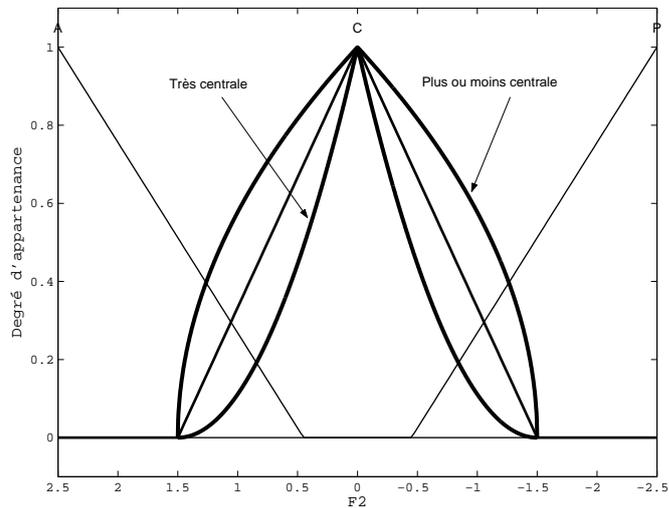


FIG. 2.4 – Illustration de la modification par les adverbess d'intensité de la forme des fonctions d'appartenances.

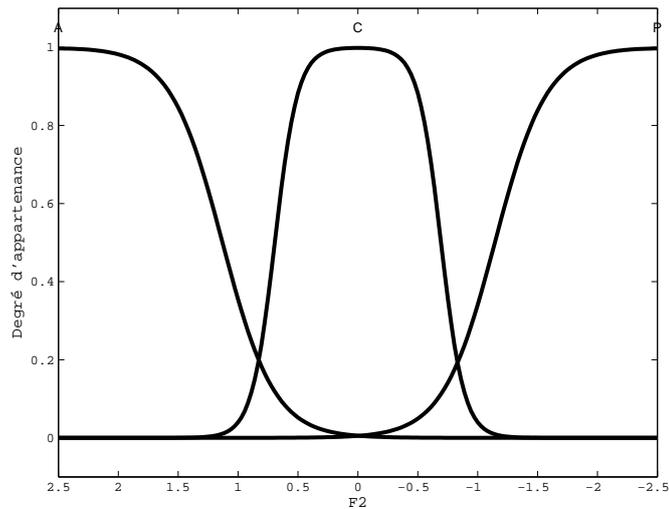


FIG. 2.5 – Fonctions d'appartenance conformes à la notion d'effet attracteur des prototypes.

sous-tend l'approche de la phonétique qui est développée dans nos travaux. Nous aurons cependant l'occasion d'y faire une brève allusion dans la Section 7.6 pour modéliser le degré d'appartenance du locuteur à un dialecte, ainsi que dans la Section 7.7, pour poser les bases d'un système expert flou susceptible d'améliorer les scores de classification et d'envisager cette classification sous un angle plus linguistique.

Il est en outre légitime de se demander si les dialectes constituent des sous-ensembles flous, et si, par exemple, la forme de ces sous-ensembles obéit au principe attracteur du prototype. Cela revient à se poser la question de savoir si les locuteurs possèdent des degrés d'appartenance à un dialecte, si les catégories les plus proches se chevauchent (au fur à mesure qu'un locuteur s'éloigne du prototype de la catégorie A, se rapproche-t-il de la catégorie B ?), etc. Nous fournissons une réponse très sommaire dans la section suivante, et nous opterons, sur des bases très hypothétiques, en faveur d'une représentation floue dans la Section 7.7. Une réponse plus empirique pourra être obtenue par le biais d'expériences de perception s'inscrivant dans le prolongement de nos travaux.

### 2.3.4 La notion de catégorie dialectale

Dans cette Section, nous abordons la question du nombre de classes dialectales et de leur forme à travers les études de perception, de contact linguistiques et de classification automatique.

Daniels ([Dan90], 27) rappelle l'aspect artificiel que comporte une étude d'identification des accents par des sujets humains :

Clearly, not all native speakers of English are interested in regional accents, so that the attempt to identify consciously a speaker's regional accent is not the everyday objective of a listener.

À partir de 40 échantillons d'accents anglais du monde entier d'une durée moyenne de 7 secondes, les 30 auditeurs anglophones natifs de l'étude de Daniels (majoritairement britanniques) sont capables, en moyenne, d'identifier avec précision environ 11 accents. Il était également demandé aux sujets de spécifier quels éléments leur avaient permis de reconnaître ces accents. Leurs réponses font apparaître deux types de critères : un trait linguistique ou phonétique particulier et la ressemblance avec l'accent d'un individu qu'ils connaissent, qu'il s'agisse d'une connaissance directe ou d'une célébrité. Parmi les traits phonétiques les plus mentionnés (qualité des voyelles et des consonnes, etc.) 15 % des réponses d'ordre linguistique concernent la rhoticité ou la réalisation du /r/. La Table 2.3 récapitule les accents des Îles Britanniques de l'étude de Daniels du mieux au plus difficilement identifié ; le score correspond au nombre d'identifications correctes. Dans ce tableau, les accents apparaissant sur la même ligne sont æquo ; pour des raisons d'homogénéité de la présentation, la terminologie anglaise a été conservée.

Malheureusement, la littérature qui traite de la perception des dialectes ne s'est, à notre connaissance, que rarement penchée sur la manière dont sont organisées les catégories dialectales dans l'esprit du locuteur natif. En 2004, Clopper et Pisoni ([CP04]) passent en revue les études marquantes dans le domaine de la perception du dialecte. Ces études abordent l'attitude des sujets vis-à-vis d'un échantillon de parole dialectale, leur capacité, parfois limitée, à identifier cet échantillon comme provenant d'une zone géographique

TAB. 2.3 – Accents des Îles Britanniques en fonction de leur rang d’identification.

Rang	Score sur 30	accent
1	28	RP, Cockney
2	23	(Adoptive) RP
3	19	Newcastle
4	14	South Wales
5	13	Leeds
6	11	Dublin
7	9	Glasgow
8	8	Belfast, Liverpool
9	7	North Yorkshire
10	6	Devon, Durham
11	5	North Wales
12	4	Norfolk, Lancashire
13	3	Cornwall, Birmingham, Black Country
14	2	Aberdeen, Ayrshire, Southern Ireland, Leicestershire, Bristol
15	1	Somerset
16	0	Hampshire, Oxford, Worcestershire

déterminée, ou encore les corrélats acoustiques des indices utilisés par les sujets dans leur tâche de classification. Aucune ne semble aborder la question de l’organisation des classes, qui est pourtant déterminante, ne serait-ce que pour pouvoir véritablement utiliser des termes tels que classes, classification ou catégories. L’étude de Clopper et Pisoni apporte un début de réponse concernant le nombre de classes que des sujets naïfs sont capables d’identifier : en anglais américain, il semblerait que trois régions dialectales, la Nouvelle Angleterre, le sud et l’ouest, puissent être identifiées de façon très fiable. Dans leur tâche de catégorisation avec un choix forcé à six classes, le taux de classification correcte est, certes, au-dessus du hasard, mais néanmoins très faible. Clopper et collègues ([CLP06], 567) rapportent, à partir d’une tâche de classification libre, que des sujets naïfs sont capables de distinguer 4 variétés régionales. Il est donc possible d’inférer que le nombre de catégories dialectales identifiables en anglais américain se situe entre 3 et 6. Cependant, les auteurs montrent ([CLP06], 569) que le taux de classification correcte pour un choix forcé avec 4 classes est d’environ 42 %, ce qui n’est pas très élevé.

Un concept intéressant des études de Clopper et collègues est celui de la saillance d’un

dialecte (notre traduction de « markedness » dans ce contexte). Clopper *et al.* ([CLP06], 567) le définissent ainsi :

markedness refers to the extent to which a given dialect has unique characteristics that differentiate it from [...] other dialects...

Les auteurs ajoutent que certains dialectes sont plus saillants que d'autres, et semblent définir cette saillance comme une fonction du nombre de traits phonétiques dont un dialecte a l'exclusivité. À partir de jugements de similitude entre locuteurs de 4 dialectes présentés par paires, [CLP06] concluent, par le biais du Multidimensional Scaling, que les deux dimensions les plus pertinentes utilisées dans les jugements de similitude sont la géographie (partition nord/sud) et la saillance. S'il était avéré – ce dont nous ne doutons pas – que ce concept puisse être transposé aux Îles Britanniques, ceci constituerait une piste pour la suite de nos travaux.

Les études sur les contacts entre dialectes offrent une ébauche de réponse à la question de la forme et du nombre de classes dialectales qu'il est possible de distinguer sur les Îles Britanniques. Le concept qui aborde indirectement ce thème est celui du nivellement dialectal (« dialect levelling »). Ce phénomène, souvent mentionné dans la littérature ([Bri02, Ker01, Ker03]), fait référence à l'érosion des formes dialectales marquées ; il peut conduire au regroupement de plusieurs variétés locales « minoritaires » en une seule. Certains considèrent d'ailleurs que « Estuary English » est une variété issue d'un nivellement ([FD99], 11, [Bri02], 63). Est-ce à dire qu'entre l'époque des données représentées dans le *Linguistic Atlas of England* ([OSW78]) et aujourd'hui, le nombre de classes qu'il est possible de mettre en évidence a diminué ? Kerswill apporte un début de réponse ([Ker01], 45) :

over one or more generations, families have abandoned these [traditional rural] dialects [and] there are fewer differences between ways of speaking in different parts of the country...

Ce commentaire peut-être interprété de deux manières : soit le nombre de classes reste le même mais la séparation entre classes repose sur des indices de plus en plus subtils ; soit le nombre de classes diminue effectivement. Le nivellement a été accompagné, dans un deuxième temps, par l'émergence de variétés urbaines très distinctes. Ces variétés urbaines subissent à leur tour un nivellement qui semble se manifester, pour Kerswill ([Ker01]),

par le biais d'une standardisation. On pourrait donc croire dans un premier temps que la bonne option consiste à admettre que le nombre de classes diminue. Cependant, la formation d'une nouvelle variété passe par l'apparition de nouveaux traits ([FD99]); il ne s'agit donc pas d'un regroupement de plusieurs variétés locales sur le critère du grand dénominateur de traits communs. Il ne s'agit pas non plus de l'adoption de la variété de prestige par toute la population ([Bri02]). Les études que nous avons consultées sur les contacts linguistiques, quoique informatives, ne permettent pas non plus de répondre à la question du nombre de catégories dialectales et des relations qu'elles entretiennent.

Pour ce qui est de la classification automatique, nous rapportons deux études qui concernent les dialectes de l'anglais. Barry et collègues ([BHN89]) proposent une approche phonétique de la question des accents dans la reconnaissance automatique de la parole, à partir de 4 accents régionaux : américain (A), écossais (E), anglais du nord (N) et anglais du sud (S). Leur méthode, dont l'un des objectifs est de maîtriser l'impact de la variation acoustique inter-locuteurs, consiste à obtenir une représentation de l'espace acoustique d'un locuteur en comparant les voyelles de son système entre elles. Ne plus comparer la production d'un locuteur à des valeurs acoustiques de référence (comme la moyenne d'un groupe) permet de s'affranchir en partie du problème de la normalisation. En outre, l'approche est intéressante car elle exploite explicitement la connaissance phonologique des systèmes vocaliques des dialectes. Par exemple, dans les quatre phrases test qu'utilisent les auteurs, une très faible distance entre <two> et <cooked> révèle la présence probable de la convergence FOOT-GOOSE, typique de E, et qu'on ne retrouve dans aucune des trois autres zones. Une proximité remarquable entre <cooked> et <butter> permettrait de mettre en évidence l'absence de scission FOOT-STRUT, diagnostique de N, et excluant les trois autres variétés. On pourrait même aller jusqu'à considérer – ce que les auteurs ne font pas – qu'une distance relativement importante entre la voyelle de <puddings> (qui serait prononcé [ʌ]) et celle de <cooked> trahirait un locuteur du nord pris en flagrant délit d'hypercorrection (voir Section 3.2.1.1). Les trois premiers formants vocaliques sont ensuite extraits, transformés en ERB, et une distance euclidienne est calculée entre chaque

TAB. 2.4 – Exemple de pondération pour la classification des dialectes.

Antécédent	Conséquent			
	S	N	E	A
$d(\text{father}, \text{after}) > d(\text{cat}, \text{after})$	-1	1	0	1
$d(\text{father}, \text{after}) < d(\text{cat}, \text{after})$	1	-1	0	-1
$d(\text{father}, \text{after}) = d(\text{cat}, \text{after})$	-1	0	1	-1

voyelle. Les comparaisons de distances subséquentes sont effectuées entre deux paires de voyelles ayant une voyelle en commun ; l’une des deux paires comporte deux phonèmes différents quel que soit le dialecte. Un système de pondération trivalent à partir de règles est ensuite mis en place : -1 indique que le trait phonético-phonologique en question joue en défaveur du dialecte concerné, 0 ne fait pencher la balance ni dans un sens ni dans l’autre, et 1 indique que le trait phonétique donne l’avantage au dialecte concerné. Nous avons réadapté un exemple de règle de la publication originale ([BHN89], 360) dans la Table 2.4. La première règle se lit de la façon suivante : si la distance entre la voyelle accentuée de <father> et celle de <after> est supérieure à la distance entre <cat> et <after>, alors, le poids en faveur de l’anglais du sud est de -1, etc. En effet, le système de l’anglais du sud n’a qu’un phonème pour PALM et BATH – i.e. <father> et <after> sont phonologiquement identiques – et deux phonèmes pour TRAP et BATH – <cat> et <after> sont phonologiquement différents. Or, si l’antécédent de cette première règle est vrai, alors le système vocalique concerné a vraisemblablement un seul phonème pour TRAP et BATH, ce qui n’est pas le cas de l’anglais du sud, mais correspond, par exemple, aux variétés du nord de l’Angleterre.

Sur les 58 locuteurs testés dans la tâche de classification (11 A, 6 E, 19 S et 20 N + une locutrice d’Irlande du Nord et un locuteur australien), 43 (73 %) sont correctement classés, 9 ont été classés dans le mauvaise catégorie et 4 avaient des scores ex æquo. Parmi les caractéristiques de cette classification, on retiendra que certains locuteurs du nord de l’Angleterre avaient une distinction entre FOOT et STRUT, ce qui a influencé le taux de classification correcte pour N (13/20). Ce dernier point, que nous aborderons à la lumière

de nos données (Sections 7.5 et 7.7), est fondamental dans l'attribution d'une classe à un locuteur. En effet, faut-il considérer qu'un locuteur du nord de l'Angleterre qui a appris la distinction FOOT-STRUT se situe quelque part entre le pôle accent du nord typique/accent du sud typique ? Est-il possible, instructif ou souhaitable de lui attribuer un degré d'appartenance à chacun de ces pôles ? Comment quantifier d'ailleurs l'appartenance à ces pôles ? Une mesure telle que la différence entre le nombre de caractéristiques phonétiques que le locuteur partage avec chacun des pôles n'est pas adaptée puisque, en raison de degrés de stigmatisation variant selon le trait (l'absence de scission FOOT-STRUT est très marquée), toutes les caractéristiques n'ont pas le même impact.

Huckvale ([Huc04]) utilise la même idée que Barry et collègues ([BHN89]), selon laquelle les distances entre les voyelles d'un même locuteur constituent une représentation robuste de son espace acoustique, et surtout de son système. Il n'utilise cependant pas des règles explicites pour la pondération des distances entre voyelles comme celles rapportées dans la Table 2.4. Nous décrivons la méthode de calcul des distances ACCDIST dans la Section 7.2 car c'est celle que nous utilisons également. Dans la procédure, seuls les paramètres acoustiques diffèrent entre l'étude de Huckvale et la tâche de classification automatique que nous décrivons dans la Section 7.2. Ce dernier utilise les 20 phrases du corpus *Accents of the British Isles* ([DRBT04]; nous employons le même corpus, mais une sous-partie différente). Les voyelles accentuées sont coupées en deux et des paramètres sont calculés sur chaque moitié. À partir des mesures des 4 premiers formants, Huckvale obtient des scores de classification correcte des locuteurs en dialectes entre environ 50 et 60 %, pour les conditions « tous sexes » et « même sexe », respectivement ; une dégradation importante (score de 35 %) intervenant dans la condition « autre sexe », où le locuteur testé et le modèle sont de sexes différents. L'utilisation de formants centrés-réduits (par locuteur) fait passer les scores à environ 72 et 73 %, la condition « autre sexe » ne donnant que 59 %. Il apparaît donc que la normalisation des formants, quoique importante dans l'amélioration des scores des conditions « tous sexes » et « même sexe », ne règle pas la question de la variation entre hommes et femmes. À partir de mesures de l'enveloppe

spectrale obtenues par le biais d'un banc de filtre auditif à 19 canaux<sup>49</sup>, la méthode ACC-DIST (Section 7.2) est utilisée pour la classification. Les taux de classification correcte pour les conditions tous, même et autre sexe sont de 86,9 %, 87,2 % et 81,4 %. Cette étude démontre qu'il est possible d'obtenir une représentation spectrale de l'espace vocalique d'un locuteur qui préserve l'information dialectale tout en étant relativement insensible aux caractéristiques non phonétiques propres à ce locuteur (e.g. le sexe).

Définir un nombre de classes dialectales sur les Îles Britanniques *a priori* semble donc être une entreprise hasardeuse. L'étude de Daniels ([Dan90]) fait apparaître des taux d'identification faibles (Table 2.3); en effet, sur les 28 dialectes britanniques de l'étude, seuls 4 ont été identifiés avec précision par plus de la moitié des sujets. Il semble évident qu'une tâche d'identification de 40 accents du monde entier présente un haut niveau de difficulté; un paradigme de choix forcé révélerait très probablement une représentation des dialectes plus fine par les sujets. Les études ayant trait aux contacts linguistiques n'apportent pas de réponse tranchée quant au nombre de classes; elles mettent cependant en avant une instabilité des catégories dialectales qu'il convenait de mentionner. Enfin, les deux études de classification automatique ([BHN89, Huc04]) décrites prouvent que la machine peut classer de façon satisfaisante jusqu'à 14 dialectes; rappelons néanmoins que, comme c'est le cas dans nos travaux, la méthode présente certes l'avantage d'inclure la connaissance phonologique, mais elle reste dépendante du texte.

## 2.4 Cadre statistique

Nos travaux de recherche nous ont conduit à accorder une place prépondérante aux méthodes d'analyse quantitative. Au fil de notre recherche, le bien-fondé d'une dichotomie entre études expérimentales et études observationnelles s'est imposé de manière évidente (voir [vB02]). L'emploi de l'inférence statistique, domaine regroupant des tests

---

<sup>49</sup>Chaque voyelle est définie par un vecteur de 40 paramètres : 19 paramètres pour chaque moitié d'une voyelle; ces données sont réduites par soustraction de la moyenne, et celle-ci est ajoutée comme paramètre supplémentaire.

d'hypothèse comme l'analyse de la variance (ANOVA), suppose l'élaboration préalable d'un plan d'expérience extrêmement minutieux visant à contrôler l'effet des facteurs que l'on ne souhaite pas prendre en compte. De plus, ces tests présupposent un caractère aléatoire, objectif qu'il n'est possible d'atteindre que par le biais de la mise en œuvre de techniques d'échantillonnage éprouvées, et ce critère n'est que très rarement satisfait en phonétique acoustique. En effet, la phonétique acoustique se place par nature dans le domaine de l'observation. Nous nous sommes donc tourné vers les méthodes descriptives, qui offrent bien plus de souplesse quant à leurs réquisits. Les diverses techniques utilisées sont brièvement présentées ci-après. Nous n'aurons recours au formalisme mathématique que lorsque celui-ci nous paraît indispensable.

### 2.4.1 Les mesures de distance

Calculer une distance dans un espace unidimensionnel, et *a fortiori* multidimensionnel, ne va pas de soi ; il convient que cette distance ait un sens. Ceci n'est pas anodin : par exemple, nous avons vu (Section 2.3.1) que calculer l'écart absolu en Hz entre deux fréquences ne correspond pas à la manière dont l'être humain perçoit la hauteur. Il est donc légitime de soulever la question d'une éventuelle transformation des variables et de la distorsion engendrée par la réduction de  $p$  paramètres à un score composite censé représenter la distance entre deux objets. Nous utiliserons principalement la distance dite « euclidienne », donnée dans l'Équation 2.6, où  $x_{ik} - x_{jk}$  est la différence entre les points  $x_i$  et  $x_j$  mesurée sur le paramètre numéro  $k$ . Cette distance est un cas particulier de la distance de Minkowski (Équation 2.7). En pratique, nous nous sommes limité aux cas où  $r = 2$  (distance euclidienne) et  $r = 1$ , distance de Manhattan (ou city block, ou encore taxicab). D'autres métriques existent ; un grand nombre d'entre elles est répertorié dans [ELL01] et [GL86].

$$d_{ij}^2 = \left( \sum_{k=1}^p (x_i^k - x_j^k)^2 \right)^{\frac{1}{2}} \quad (2.6)$$

$$d_{ij}^r = \left( \sum_{k=1}^p |x_i^k - x_j^k|^r \right)^{\frac{1}{r}} \quad (2.7)$$

Si le choix d'une mesure de distance est primordial, et est déterminé *in fine* par le type de données spécifique dont on dispose, le problème connexe de la pondération des variables (des dimensions ou paramètres) l'est tout autant. La distance de Minkowski attribue de fait une pondération équivalente à chacune des variables d'origine ; par conséquent, cette mesure est sensible aux différences d'échelles. En outre, plus le nombre  $r$  de l'Équation 2.7 est élevé, plus important est le poids affecté aux variables dont la différence absolue entre deux objets est grande. Le fait de centrer-réduire les variables originales revient à leur accorder un poids identique implicitement ; mais il est également possible de pondérer explicitement. À ce propos, [ELL01], 48-49, notent qu'on peut considérer que la pertinence d'une variable est inversement proportionnelle à l'amplitude de sa variation. En réalité, cette approche risque de masquer la variation inter-classe, et c'est précisément cette dernière qui devrait être pondérée à la hausse dans une tâche de classification. Une solution consiste, surtout pour une tâche supervisée, à donner plus de poids aux dimensions pour lesquelles le rapport de la variation inter-classe sur la variation intra-classe est maximal. Dans l'Équation 2.8, la pondération  $w$  est incluse dans le calcul de la distance de Minkowski.

$$d_{ij}^{rw} = \left( \sum_{k=1}^p w_k |x_i^k - x_j^k|^r \right)^{\frac{1}{r}} \quad (2.8)$$

Nous avons par exemple tenté de déterminer empiriquement les poids relatifs de la durée et des paramètres spectraux dans [FP06] : en utilisant la méthode décrite dans la Section 7.2, nous avons fait varier un facteur de pondération  $w$  de 0 à 1 pour la durée, et avons multiplié à chaque itération les paramètres spectraux par  $1 - w$ . La pondération optimale correspondait au taux de classification le plus élevé. Mentionnons également le programme Optimal Variable Weighting ([ML01]). Ce programme s'efforce de trouver des poids optimaux pour chaque variable afin que la distance entre classes soit maximale, et

celle à l'intérieur des classes, minimale. Ce programme a été utilisé sans succès sur nos données, nous nous en tiendrons donc à cette rapide allusion.

## 2.4.2 La classification hiérarchique

La classification hiérarchique est une technique qui vise à obtenir une représentation arborée, le dendrogramme, ou arbre hiérarchique, d'un ensemble d'individus. Nous avons utilisé cette technique dans la Section 7.3 ; les arbres représentés sont binaires en cela que les individus  $y$  sont regroupés deux à deux. Détaillons à présent la méthode employée. Dans un premier temps, la procédure consiste à obtenir les distances entre les individus à classer pris deux à deux ; ce point vient d'être abordé dans la Section 2.4.1. Puis, à partir de la partition de départ, où chaque individu constitue une classe, les deux individus les plus proches sont réunis. Les distances entre classes sont alors recalculées. C'est ici qu'intervient la notion de critère d'agrégation ; la distance entre la classe qui vient d'être formée et les autres classes peut être exprimée de différentes façons (un exemple très clair du fonctionnement des méthodes agglomératives est donné dans [ELL01], 57-58). Nous avons déterminé empiriquement que la méthode du lien moyen<sup>50</sup> impliquait la distorsion moyenne la plus faible entre les distances originales et les distances telles qu'elles sont représentées dans l'arbre. La méthode du lien moyen consiste à définir la distance entre deux classes comme la moyenne des distances entre chaque paire d'individus n'appartenant pas à la même classe. L'opération est ainsi répétée jusqu'à ce que toutes les classes soient réunies en une seule (pour d'autres méthodes, voir [ELL01] 55-89). Un diagnostic de la distorsion entre la matrice de distances originale et l'arbre final peut être obtenu au moyen du coefficient de corrélation cophénétique. Il consiste à calculer une corrélation linéaire classique entre la matrice de distances originales et la matrice cophénétique, i.e. la matrice des distances entre individus deux à deux, ces distances étant données par la hauteur à laquelle deux individus sont regroupés dans l'arbre.

---

<sup>50</sup> « Average linkage » ou encore « unweighted pair-group method using the average ».

### 2.4.3 L'analyse en composantes principales

S'il est impossible de représenter graphiquement des individus (voyelles, locuteurs, etc.) dans un espace à plus de deux ou trois dimensions, ceci ne doit en aucun cas constituer un frein à l'inclusion d'autres variables. D'ailleurs, il est toujours possible de représenter ces individus par leurs coordonnées sur toutes les dimensions prises deux à deux (ce que certains logiciels nomment « scatterplot matrix ». Voir, par exemple, la Figure 5.6 dans la Section 5.4.). Ceci constitue un excellent moyen de visualiser les variables corrélées ; mais en représentant autant de nuages de points sur des graphes bidimensionnels qu'il y a de combinaisons de variables prises deux à deux, la figure finale, quoique instructive, n'est pas synthétique. En effet, outre la taille importante de cette figure, les variables corrélées fournissent une information redondante (puisque en partie prévisible à partir des autres variables) car une multiplicité de variables n'obéissent souvent qu'à un nombre restreint de principes sous-jacents. L'analyse en composantes principales permet de réduire le nombre de variables initiales (corrélées) à de nouvelles variables orthogonales dans un espace réduit. On calcule les vecteurs propres et valeurs propres de la matrice de covariance des paramètres centrés-réduits. Ces vecteurs propres, orthogonaux, sont ensuite ordonnés en fonction de la valeur propre qui est associée à chacun. Cette valeur traduit la proportion de variation dans l'espace des paramètres originaux qui est conservée si on projette les données sur le vecteur propre correspondant. La première composante principale est définie par le vecteur propre associé à la valeur propre la plus élevée, et ainsi de suite. Pour plus de détails, voir [WFH86], 273-290, [Sap90], 159-186, [Lag98], *passim*.

### 2.4.4 L'analyse linéaire discriminante

L'analyse linéaire discriminante est une technique de classification avec apprentissage supervisé. Là où l'analyse en composantes principales vise à décrire au mieux la dispersion des données dans un espace à dimension réduite (sans tenir compte de leur éventuelle organisation en classes), l'analyse discriminante consiste à déterminer le meilleur hyper-

plan séparant les différentes classes d'individus. Pour ce faire, on cherche à maximiser le rapport de la variation inter-classe (e.g. la matrice de covariance calculée sur l'ensemble des données) sur la variation intra-classe (e.g. la moyenne des matrices de covariances de chaque classe). Les données originales sont alors projetées dans l'espace des vecteurs propres contenant le plus d'information discriminante. Pour un aperçu de la méthode et des détails plus techniques, voir [Lag98], 113-125, [Sap90], 403-428, [DHS01], 114-124 et 215-281, pour une approche plus exhaustive des fonctions discriminantes ; dans ce cadre plus général, voir également [DMS<sup>+</sup>02].

### 2.4.5 « K-means clustering » et qualité d'une partition

Le « k-means clustering » est une méthode de classification non supervisée ([DHS01], 526-528, [Nab02], 101-104) qui, dans sa forme élémentaire, consiste à localiser  $c$  vecteurs moyens (ou centres de classes) –  $c$  étant déterminé à l'avance – dans un espace multidimensionnel, pour lesquels la somme des distances de chaque individu au centre le plus proche est minimale. La fonction `kmeans` du logiciel Matlab prévoit une première phase pendant laquelle, à chaque itération, les points sont associés au centre le plus proche, qui est ensuite recalculé jusqu'à ce que les coordonnées des centres soient relativement stables. Dans un second temps, à chaque itération, chaque point est associé à un centre différent si cela permet une diminution de la somme des distances, et les centres sont mis à jour.

Parmi les nombreuses techniques permettant de juger la qualité d'une partition ([ELL01, CH06]), nous avons utilisé la valeur de silhouette, dont le calcul est donné dans l'Équation 2.9.  $s_j$  est la valeur de silhouette de l'objet  $j$ ,  $a_{pj}$  représente la distance moyenne de l'objet  $j$  aux autres objets appartenant à la classe  $p$ . Si  $d_{qj}$  est la distance moyenne entre l'objet  $j$  et tous les objets appartenant à une classe  $q$ ,  $q \neq p$ ,  $b_{pj}$  est la valeur  $d_{qj}$  minimale calculée pour  $q = 1 \dots c$ ,  $q \neq p$  ( $c$  étant le nombre de classes). Autrement dit,  $b_{pj}$  mesure la dissimilarité de l'objet  $j$  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  $j = 1 \dots N$  ; 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. C'est ce que nous avons fait, par exemple, dans la Section 5.6.

$$s_j = \frac{b_{pj} - a_{pj}}{\max\{a_{pj}, b_{pj}\}} \quad (2.9)$$

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 points qui ne peuvent pas être clairement attribués à une classe, et une valeur proche de  $-1$  indique que le point en question n'appartient vraisemblablement pas à la bonne classe. Everitt *et al.* ([ELL01], 105) ajoutent qu'une valeur de silhouette moyenne inférieure à  $0,2$  révèle une absence totale de structure en classes, et une valeur supérieure à  $0,5$  traduit une bonne classification.

## 2.4.6 La régression

Nous avons employé la technique de la régression dans le but de décrire la trajectoire des formants des voyelles. La régression permet d'exprimer une relation entre plusieurs variables. Dans notre cas, on recherche les paramètres d'une fonction affine (Équation 2.10) passant, pour chaque voyelle, au plus près des valeurs formantiques dans le plan temps (variable indépendante)/fréquence (variable dépendante).

$$f(x) = a + bx \quad (2.10)$$

« Passer au plus près » se décline en deux versions : calcul des moindres carrés ordinaires et calcul des moindres carrés pondérés. Le premier attribue un poids équivalent à chaque point, le second pondère à la baisse les points apparemment déviants. Le calcul de la pente,  $b$ , par la méthode des moindres carrés ordinaires est donné dans l'Équation 2.11, où  $n$ , dans notre cas, est le nombre de valeurs formantiques,  $x$  est la valeur de temps et  $y$ , la valeur fréquentielle.

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.11)$$

La régression utilisant les moindres carrés pondérés est une procédure itérative donnant une estimation robuste à la présence de valeurs déviantes. On peut considérer que les moindres carrés ordinaires sont un cas particulier des moindres carrés pondérés où le poids de chaque point est équivalent. L'algorithme estime dans un premier temps la droite des moindres carrés (Équation 2.11), puis, l'écart entre chaque point observé et cette droite est mesuré, et chaque point observé reçoit un poids inversement proportionnel à sa distance avec la droite. Une nouvelle droite est calculée en tenant compte de ces poids, et ainsi de suite jusqu'à ce qu'il n'y ait plus de modification importante de la droite (voir [WW91]). La fonction de Matlab utilisée pour ce type de régression est `robustfit`; et pour la régression simple, `regress`.

Comme nous le verrons dans la Section 6.2, les formants peuvent décrire une trajectoire courbe, auquel cas ajuster une droite aux points dans l'espace temps/fréquence devient aberrant. Dans ces cas-là, l'estimation des paramètres d'une fonction polynôme sera réalisée avec la fonction `polyfit` du logiciel Matlab. L'Équation 2.12 donne la forme d'une fonction polynôme, où  $n$  est le degré du polynôme et  $b_i \dots_n$ , les coefficients. L'estimation de ces coefficients se fait, comme pour la régression ordinaire, avec le critère des moindres carrés. Ce qu'il est, en pratique, important de retenir pour nos besoins, c'est qu'un polynôme d'ordre  $n$  ne peut pas générer une courbe comportant plus de  $n - 2$  points d'inflexion.

$$f(x) = a + b_1x + b_2x^2 + \dots + b_nx^n \quad (2.12)$$

Les autres techniques utilisées dans ce travail seront présentées au fur et à mesure que nous les rencontrerons.

Après avoir introduit dans ce chapitre les notions fondamentales de notre étude, nous passons en revue les caractéristiques phonético-phonologiques des dialectes de l'anglais des Îles Britanniques dans le Chapitre 3.



# Chapitre 3

## Situation géolinguistique et anglais standard

Ce chapitre a pour but de faire le point sur les connaissances actuelles concernant les aspects phonético-phonologiques des dialectes des Îles Britanniques. Nous nous concentrons sur le domaine segmental ; certaines caractéristiques suprasegmentales seront néanmoins abordées plus bas dans le Chapitre 5. Dans ce qui suit, les particularités des systèmes vocaliques constituent l'essentiel du propos. Cependant, bien que nous n'abordions pas le cas des consonnes dans notre étude empirique (Partie II), nous avons tenu à mentionner les traits consonantiques les plus saillants des zones géographiques passées en revue. Ce chapitre s'articule en deux temps : la Section 3.1 traite de la notion d'anglais standard et fournit une description détaillée du système vocalique de cette variété ; et la Section 3.2 dresse un inventaire des traits phonético-phonologiques diagnostiques des principales zones dialectales des Îles Britanniques.

### 3.1 L'anglais standard

Une description exhaustive de l'histoire de l'anglais standard et une rétrospective des points de vue des experts sur la question seraient pour le moins rebutantes. Nous nous

contenterons de définir ce que recouvre ce concept et de fournir une description phonétophonologique détaillée de cette variété. Ce dernier volet est capital puisque c'est à l'aune de cette description que seront examinés les dialectes de notre corpus.

Le terme « Received Pronunciation » (RP) a longtemps été employé, notamment dans le domaine de l'enseignement de l'anglais comme langue étrangère. Il désigne sans ambiguïté un type de prononciation (et ne fait donc aucune référence au lexique ou à la syntaxe), qui, dans sa forme la plus conservatrice, serait à peine parlée par 3 à 5 % de la population anglaise ([HTW05], 3). À notre sens, le terme est très lourdement connoté puisque, bien que Wells ([Wel90]) tente de l'associer à un effort descriptiviste, c'est bien une norme prescrite qui est évoquée par RP. Ce terme a été employé pour la première fois en 1869 par le philologue Alexander Ellis, puis remis au goût du jour en 1926 par Daniel Jones, figure emblématique de la phonétique moderne ([Upt04], 217). Ce concept a été désigné depuis le début du 20<sup>e</sup> siècle par les termes « Public School Pronunciation », « BBC English » (c'est ce terme qu'utilisent Roach et collègues [Jon03]), ou encore « Queen's English », entre autres, ce qui traduit admirablement le caractère hautement élitiste souvent associé à cette prononciation. C'est la raison pour laquelle nous avons abandonné le terme. Wells ([Wel82], 279) parlait de « Varieties of RP » ; c'est précisément cette acception plus large (mais cependant limitée aux variétés « natives ») que recouvre notre conception de l'anglais standard.

Lorsque l'on parle d'anglais standard<sup>51</sup> sur les Îles Britanniques, on s'empresse de signaler que cette variété n'a aucune connotation régionale, et qu'il s'agit donc d'un « sociolecte ». Certes. Mais il n'en demeure pas moins que les traits de prononciation de l'anglais standard sont typiques du sud-est de l'Angleterre. D'ailleurs, l'histoire de l'anglais standard commence dans le sud-est, comme le rappelle Honey ([Hon97], 93 *sqq.*), et Nolan ([Nol98]) considère que ce qu'il nomme « RP » ou « near-RP » est lié aux accents du sud-est<sup>52</sup>. Après avoir identifié les principaux facteurs qui ont conduit à l'émergence des

---

<sup>51</sup>Ce terme fera exclusivement référence à l'aspect phonétophonologique.

<sup>52</sup>Rappelons aussi que, dans l'étude de Evans et Iverson ([EI07]) décrite dans la Section 2.3.3, plus les voyelles d'un locuteur originaire du nord se rapprochent de l'anglais standard, plus ce locuteur est jugé proche du sud (sur un continuum nord/sud) par des phonéticiens.

variétés standards en Europe occidentale dès le 15<sup>e</sup> siècle<sup>53</sup>, l'auteur rappelle que l'anglais standard est associé aux zones de Londres, Cambridge et Oxford. Sa forme écrite trouve comme vecteur de diffusion l'imprimerie naissante<sup>54</sup> et sa forme orale trouvera un moyen de propagation bien plus tard par l'intermédiaire du réseau des « public schools ». D'après Honey ([Hon97], 94), la radio dès 1922, puis la télévision, ont pris le relais.

Dans une tentative de définition de ce que nous appellerons « anglais standard », nous donnerons, en guise de préliminaire, l'avertissement de Wells ([Wel91]) :

it is important not to lose sight of the fact that accents, and more generally varieties of a language, are not objective entities so much as mental constructs. As such, their definition and circumscription are open to considerable debate.

Parmi les définitions possibles de ce que certains nomment RP, Wells ([Wel91]) se demande, sur un ton légèrement ironique, si l'on ne ferait pas mieux de considérer qu'il ne s'agit de rien d'autre qu'une norme codifiée à l'usage des apprenants étrangers. Si cette idée est effectivement étroitement liée au concept que nous tentons de définir, il n'en demeure pas moins qu'elle ne constitue pas une définition. En effet seuls les critères présidant à l'élaboration de cette norme sont susceptibles de posséder un caractère définitoire d'une quelconque utilité pour la suite de notre propos, l'utilisation qui est faite de cette norme n'étant que secondaire ici. En d'autres termes, s'il est vrai que la description d'une norme de prononciation de l'anglais britannique a été en grande partie motivée par l'enseignement de l'anglais comme langue étrangère<sup>55</sup>, cela ne doit pas occulter le fait que le choix des traits de prononciation est motivé par une réalité linguistique déterminante dans la société britannique. Remarquons cependant que l'anglais standard auquel nous faisons référence n'est pas une norme de prestige pour toutes les Îles Britanniques ; il l'est en Angleterre, mais, concernant les autres pays, Foulkes et Docherty ([FD99], 11) observent que : « RP has little cachet for many speakers, particularly those outside of England [and is] often regarded with 'hostility' in Glasgow. »

---

<sup>53</sup>Avènement de l'imprimerie, des états-nations dotés de systèmes éducatifs au niveau national ; mobilité géographique et sociale associée à l'industrialisation et à l'urbanisation.

<sup>54</sup>Caxton installe ses presses à Westminster en 1476.

<sup>55</sup>Nous en voulons pour preuve l'habitude prise dans les ouvrages semblables à [Jon67] et [Gim80] de consacrer une place importante aux conseils pour les apprenants.

Quel que soit le terme utilisé pour désigner cette norme de prononciation, il va de soi que cette dernière est profondément instable en diachronie ; l'anglais standard de Daniel Jones n'est pas celui de John Wells. Il apparaît dès lors capital de comprendre qu'une description phonétique n'est rien d'autre qu'un instantané ; un anglais standard ne peut être défini qu'à travers l'époque pendant laquelle il a cours. Il est la marque d'une communauté, et le fait que cette communauté disparaisse, ou que ses frontières deviennent floues (comme ce fut le cas de classes sociales en Angleterre ces quarante dernières années) est lourd de conséquences.

Nous allons certes définir l'anglais standard en énumérant les traits de prononciation qui le caractérisent, mais une caractérisation plus parcimonieuse consisterait à le définir négativement ; nous dirions alors que l'anglais standard est dénué de traits de prononciation qui trahiraient l'origine géographique ou sociale d'un locuteur.

### 3.1.1 Système vocalique de l'anglais standard

Nous nous appuyerons sur plusieurs ouvrages ([Jon67], [Wel82], [Wel90], [Jon03], [Upt04]) pour une description phonétique traditionnelle du système vocalique de l'anglais standard. Les commentaires ayant trait à l'évolution acoustique des voyelles sur les quarante dernières années se fondent sur [Wel62], [Hen83], [Det97] et [HM05].

La description est structurée autour des mots-clés représentant les ensembles lexicaux popularisés par Wells ([Wel82]). L'ordre d'apparition de ces ensembles est identique à celui de Upton ([Upt04]) ; les voyelles brèves précèdent les voyelles longues, et l'exposé s'achève avec les voyelles réduites. Dans les Figures 3.1, 3.2 et 3.3, nous avons adapté les trapèzes vocaliques de [Jon03], viii-ix. Il est à noter, au vu de la présentation phonétique détaillée qui suit, que ces trapèzes présentent une prononciation assez conservatrice.

KIT a un timbre mi-fermé, proche de [ɪ], quoique plus postérieur. En position inaccentuée, cette voyelle est encore utilisée dans les morphèmes <-ed> et <-es> (certains dialectes ont [ə], voir 3.2.1.4). En revanche, [ɪ] a largement cédé la place à [ə] dans les autres contextes non-accentués où la réduction vocalique opère. Wells ([Wel91]) retrace

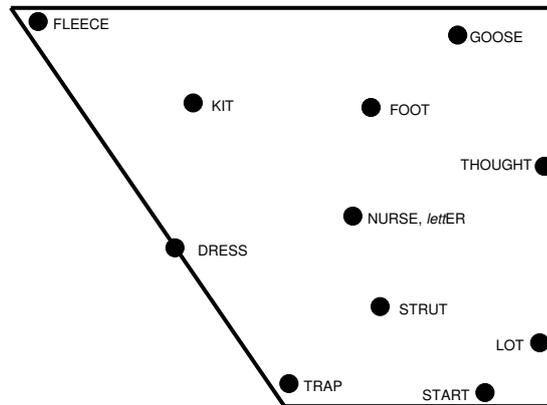


FIG. 3.1 – Monophthongues de l'anglais standard (adapté de [Jon03]).

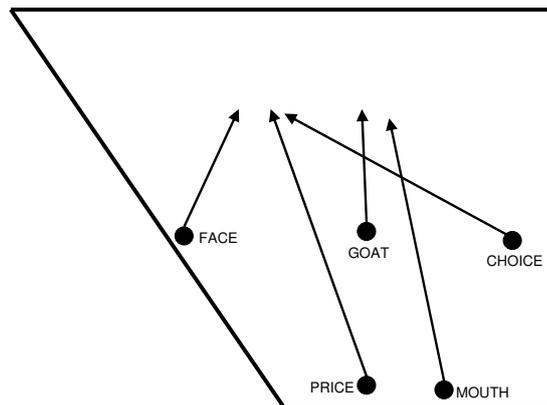


FIG. 3.2 – Diphtongues fermantes de l'anglais standard (adapté de [Jon03]).

brèvement l'histoire de l'acceptation de ce [ə] dans les descriptions de l'anglais standard : jusqu'en 1967, Jones continuait de penser que la prononciation [-ləsnəs] dans <carelessness> ne méritait pas d'être incluse dans son dictionnaire. Dix ans plus tard, dans le même dictionnaire, révisé par Gimson, ce dernier jugeait cette variante comme acceptable. Treize ans plus tard, Wells, dans son dictionnaire ([Wel90]), en fait la variante principale, reléguant [-ɪsnɪs] en seconde position. Ce choix est toujours d'actualité puisque c'est celui qui a été retenu dans la 16<sup>e</sup> édition du dictionnaire de Jones ([Jon03]). Wells ([Wel99]) a étudié ce phénomène des voyelles faibles pré-consonantiques par le biais d'un questionnaire

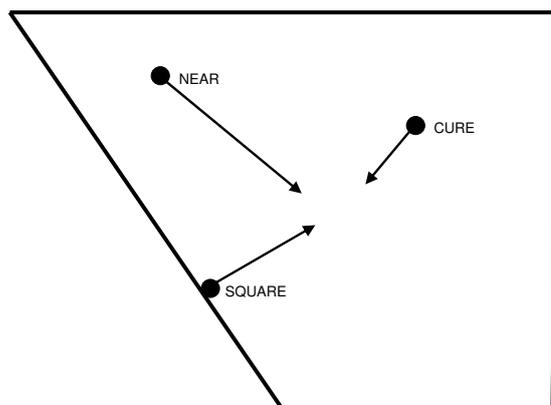


FIG. 3.3 – Diphtongues centripètes de l’anglais standard (adapté de [Jon03]).

auprès d’environ 2000 personnes. Les résultats font apparaître que 76 % des personnes interrogées préfèrent [ə]. En termes de système, on peut supposer (comme le fait [Gim80], 105) qu’une certaine pression existe encore contre une neutralisation complète de /ɪ/ et /ə/ en syllabe inaccentuée : <offices> et <officers> peuvent être distingués si les voyelles finales de ces deux mots sont prononcées respectivement /ɪ/ et /ə/. Nuancions toutefois : dans l’histoire de l’anglais, si ces risques d’homophonie ont peut-être retardé certaines neutralisations et convergences phonémiques, ils n’ont jamais constitué une entrave suffisante. Nolan ([Nol98], 80) considère que ce changement s’inscrit dans une tendance générale de réduction des voyelles non accentuées qui n’avaient pas encore schwa ; d’après l’auteur, le changement est plus avancé en anglais australien, et nous savons par ailleurs qu’une étude récente sur l’anglais américain ([FJ07]) prouve l’existence d’un contraste dans la voyelle finale des (rares) paires du type <Rosa’s> vs <roses><sup>56</sup>.

DRESS a une voyelle de type [ɛ]. Une forme plus conservatrice de l’accent présente une voyelle plus fermée. Upton ([Upt04], 222) note l’ouverture accrue relativement récente de cette voyelle, qui participe d’une tendance affectant les voyelles antérieures brèves. L’ouverture de DRESS est visible dans l’étude acoustique de Hawkins et Midgley ([HM05]) ; son évolution semble toutefois moins flagrante que celles de TRAP, FOOT et GOOSE (voir

<sup>56</sup>Les paires du type <offices>/<officers> ne permettent pas de tester ce type d’oppositions puisque la plupart des accents américains sont rhotiques.

*infra*).

La voyelle de TRAP est traditionnellement représentée par le symbole æ dans les dictionnaires de prononciation. Or il semblerait que la tendance récente à l'ouverture des voyelles antérieures brèves mentionnée plus haut soit particulièrement saillante pour celle de TRAP. Il apparaîtrait donc que la voyelle de TRAP en anglais standard ressemble de nos jours à celle des dialectes septentrionaux de l'Angleterre : [a]. Cette observation est largement corroborée par l'étude acoustique transversale de Hawkins et Midgley ([HM05]) : F1 passe de 644 Hz, pour les locuteurs âgés de 65 ans ou plus, à 917 Hz pour les 20-25 ans. Cette ouverture s'accompagne d'une postériorisation du timbre qui n'est en grande partie que le résultat de contraintes articulatoires et de propriétés acoustiques des voyelles bien connues. En effet, et ceci est particulièrement flagrant pour les voyelles antérieures, le degré de fermeture est positivement corrélé au degré d'antériorité<sup>57</sup>. Wells ([Wel82], 291-292) remarquait déjà le phénomène d'ouverture de TRAP en citant un article de journal de 1978 dans lequel un journaliste vilipendait cette tendance qu'il considérait comme une conséquence des années 1960, époque à laquelle, précise le journaliste, « working class became beautiful ». Wells poursuit son exposé en notant que cette ouverture de TRAP conduit à une convergence potentielle (et partielle chez certains locuteurs) avec la voyelle de STRUT qui, comme nous le verrons plus loin, est particulièrement ouverte en anglais standard (ceci est toujours d'actualité puisque [Upt04], 222, reprend le même exposé). Pour [Wel91], cette ouverture de la voyelle de TRAP en anglais standard peut s'expliquer comme une volonté de s'éloigner de la réalisation en Cockney, [ɛ]. Le fait que ce « flat A » s'apparente à celui des dialectes du nord n'est qu'une conséquence de ce phénomène pour [Wel91] alors que [Upt04], 222-223, semble penser que ce sont précisément les locuteurs originaires du nord qui ont fait pénétrer cette réalisation dans l'anglais standard. Fudge ([Fud77]) traite la question de la voyelle de TRAP en référence à sa propre prononciation. Il note que dans le sud de l'Angleterre et en RP, cette voyelle est parfois allongée, et tente

---

<sup>57</sup>Le trapèze vocalique actuel, qu'il soit d'inspiration articulatoire ou acoustique, présente une partie antérieure oblique, illustrant l'impossibilité physiologique de conserver le même degré d'antériorité au fur et à mesure que l'aperture s'accroît.

de détecter les facteurs de cet allongement. Cette variante allongée semble ne pas obéir à des règles bien précises, mais Fudge envisage la possibilité de deux phonèmes distincts. Le degré de formalité constitue un facteur probable ; pour l'anecdote, l'auteur offre au lecteur un moyen mnémotechnique avec les mots br[æ]ndy et sh[æ:]ndy. La longueur de la voyelle est prévisible par le degré de formalité des situations associées à la consommation de ces deux boissons (cognac et panaché, respectivement) et si cela ne suffit pas, il faut se souvenir que le brandy est un « short drink », et le shandy, un « long drink » . . .

La voyelle de LOT est postérieure, de degré d'aperture maximum (ou très légèrement fermée), et arrondie. Cette voyelle est brève, contrairement à son équivalent dans le système de l'anglais américain standard, où <dog> est volontiers prononcé [dɔ:ɡ].

Upton ([Upt04], 222) décrit la voyelle de STRUT de la manière suivante : elle est centralisée et légèrement plus ouverte que [ʌ]. Il ajoute que pour un bon nombre de locuteurs, STRUT s'apparente davantage à un [a] légèrement plus fermé et centralisé que la voyelle désignée par ce symbole ([ɐ] serait peut-être tout à fait adéquat). Nolan ([Nol98], 81) inclut cette voyelle dans les changements récents<sup>58</sup> en RP.

FOOT a une voyelle mi-fermée, antériorisée et arrondie. L'antériorisation de ce timbre est attestée dans les données de [HM05], et Henton ([Hen83], 365) notait déjà en 1983 que cette centralisation de FOOT et GOOSE était le phénomène le plus fréquemment mentionné dans la littérature ayant trait à l'évolution du système de l'anglais standard.

Traditionnellement, la voyelle de BATH est ouverte, postérieure et longue ([Jon67], 73-77). Cependant, [Upt04], 223, constate une évolution imputable à deux phénomènes distincts. D'abord, la voyelle de BATH tend à être de plus en plus brève et de plus en plus centrale sur l'axe antérieur/postérieur ; la variante longue et postérieure étant progressivement associée à un accent affecté. Ensuite, le [a] caractéristique des accents du nord semble avoir été accepté dans le système de l'anglais standard, tant et si bien, qu'il n'est pas rare d'entendre des locuteurs du nord (i.e. dont l'accent d'origine n'a pas de distinction TRAP-BATH) présentant toutes les caractéristiques de l'anglais standard utiliser [a] pour

---

<sup>58</sup>Ici, « récent » qualifie les tendances de prononciation apparues au 20<sup>e</sup> siècle.

l'ensemble BATH, voire [a] et [ɑ:] de façon interchangeable. Ceci rejoint un commentaire de Wells dans une interview au *Guardian* de 1999 faisant suite aux sondages de préférences de prononciation pour la deuxième édition du *Longman Pronouncing Dictionary* ([Byr99]) : « There is a northern pronunciation increasing among the young which the word 'chance' clearly shows ».

L'étude acoustique de Hawkins et Midgley ([HM05]) ne permet pas de commenter le cas de BATH. Des trois ensembles lexicaux susceptibles de contenir un phonème /ɑ:/ (BATH, START et PALM), seul START est présent dans les mots-test de l'étude ; or c'est précisément celui qui tend à préserver un timbre postérieur non stigmatisé.

La voyelle de CLOTH est aujourd'hui identique à celle de LOT. Dans son fameux poème pour apprenants, Trénité<sup>59</sup> avait écrit :

Mind the o of off and often  
Which may be pronounced as orphan

Cette variante ([ɔ:]) s'avère aujourd'hui totalement archaïque en anglais standard ; Upton ([Upt04]), 224 la juge même « risible ». Gimson ([Gim80], 115) explique l'abandon de la variante longue et plus fermée de ce phonème par sa similitude avec la variante utilisée en Cockney.

D'après [Upt04], 224, la voyelle de NURSE est très variable, de mi-ouverte à mi-fermée ; cette caractéristique est également mentionnée chez [Jon67], 88. La variante la plus ouverte était associée à U-RP<sup>60</sup> d'après Wells ([Wel82], 281).

FLEECE a une voyelle longue, antérieure, fermée. Elle est souvent légèrement diphtonguée : [ii]. Jones [Jon67], 65-66, conseillait aux apprenants de produire une monophthongue. Déjà en 1980, Gimson ([Gim80], 102-103) remarquait que parmi les locuteurs de RP, une légère diphtongue était plus courante qu'une véritable monophthongue. Il ajoutait que produire une monophthongue lorsque la voyelle de FLEECE était en position finale relevait d'une prononciation « over-cultivated ». Il recommandait aux apprenants de tenter de produire une diphtongue, mais seulement avec précaution.

<sup>59</sup><http://www.hep.wisc.edu/~jnb/charivarius.html> ; site consulté le 8 avril 2008.

<sup>60</sup>Une abréviation fréquente pour « upper » ou « upper-crust » RP.

La voyelle de FACE en anglais standard est une diphtongue fermante dont les points de départ et d'arrivée sont respectivement [e] et [ɪ]. Le timbre de la première partie de la diphtongue varie entre [ɛ] et [e] ([Gim80], 129). La limite de fermeture est imposée sur le plan articulaire; la limite d'ouverture est, elle, sociolinguistiquement déterminée : un premier élément de type [æ] ou [a] est typique de certains accents populaires, notamment le Cockney. Un mouvement articulaire ample étant, comme nous venons de le voir, associé à des variétés non prestigieuses, certains locuteurs marquent leur distance avec ces variétés stigmatisées en utilisant un premier élément très fermé, occasionnant un mouvement articulaire restreint.

La voyelle de PALM est longue, ouverte et quelque peu centralisée. Plus elle est postérieure, plus elle évoque ce que Upton [Upt04], 224, nomme « Refined RP ». Cet ensemble, décrit par Wells ([Wel82], 142-144), est, aux dires de ce dernier, plutôt difficile à cerner avec précision. Il comprend peu de mots courants (e.g. *father*, *tomato*, etc.) et est essentiellement constitué d'emprunts récents à d'autres langues, le timbre postérieur soulignant le caractère exotique du mot en question (e.g. *sonata*, *rajah*, *façade*, etc.).

La voyelle de THOUGHT est généralement notée [ɔ:] ([Wel90, Jon03, Upt04, OS06]). Elle s'est fermée depuis le début du 20<sup>e</sup> siècle; Nolan ([Nol98], 82) considère qu'il s'agit, dans une certaine mesure, d'un prolongement du Grand Changement Vocalique; il note [ɔ̟:].

GOAT a une diphtongue de type [əʊ]. La variante traditionnelle et rare, [oʊ], était naguère la norme. La variante [ɛʊ] appartient à une forme d'accent raffiné qui confine à la caricature. Nolan ([Nol98], 82) semble faire de cette dernière variante une question de génération :

‘no joke’ as spoken by some younger speakers seems to old speakers to sound more like ‘nay Jake’<sup>61</sup>.

La voyelle de GOOSE est souvent décrite comme voyelle postérieure longue et fermée. En réalité, un timbre antériorisé et une diphtongue de type [ʊɪ] sont très répandus, le

---

<sup>61</sup>On remarque, en effet, que chez certains locuteurs le deuxième élément de la diphtongue est au moins partiellement délabialisé : [ɛɪ] tendant vers [ɛi].

degré d'antériorisation étant négativement corrélé à l'âge des locuteurs. En examinant les relevés de formants de l'étude de Hawkins et Midgley ([HM05]), où les locuteurs sont divisés en classes d'âge, on relève immédiatement que GOOSE et FOOT ont connu une antériorisation flagrante, notamment chez les locuteurs de 20 à 40 ans. En effet, les valeurs moyennes de F2 pour GOOSE passent de 994 Hz chez les 65 ans et plus à 1616 Hz chez les 20-25 ans. Cette dernière valeur correspond, si l'on accepte le parallèle entre trapèze vocale articulaire et plan F1/F2, à un degré d'antériorité supérieur à la voyelle de TRAP. Pour FOOT, la tendance est identique, à ceci près que, dans le cas de GOOSE, l'évolution de F2 en fonction de l'âge semble plus graduelle, alors que pour FOOT, l'antériorisation est soudaine et spécifique aux 20-25 ans. Gardons-nous cependant d'en tirer des conclusions trop hâtives, chaque classe d'âge dans l'étude citée n'étant constituée que de 5 locuteurs. Gimson ([Gim80], 122) identifie deux facteurs, l'un systémique, l'autre sociolinguistique, influant sur la position de la voyelle de GOOSE en anglais standard. L'absence de phonème dans la zone de [y] permet à GOOSE de migrer vers la partie antérieure du système. En revanche, une antériorisation trop marquée est associée aux dialectes de la région de Londres, notamment populaires ; ceci jugule dans une certaine mesure une potentielle antériorisation supplémentaire en anglais standard. Sur le plan de la perception, Harrington et collègues ([HKR07]) ont montré, à partir de stimuli sur un continuum /i/-/u/, que la frontière entre les catégories /i/ et /u/ dépendait de l'âge des auditeurs. Plus précisément, les sujets les plus jeunes catégorisent dans /u/ des stimuli avec un F2 élevé (i.e. proches de [i]) que les plus âgés classent comme /i/ ; il y a donc, compte tenu notamment des résultats de Hawkins et Midgley, une symétrie entre production et perception. Les auteurs démontrent en outre une absence de compensation perceptive des effets de coarticulation (e.g. l'antériorisation après [j] n'est plus perçue comme une conséquence de la coarticulation) dans le groupe des sujets jeunes, qui pourrait être une des causes de l'antériorisation de la voyelle de GOOSE en anglais standard.

L'accent le plus traditionnel a [aɪ] pour la voyelle de PRICE. Upton ([Upt04], 225) et Olausson et Sangster ([OS06]) trouvent la transcription suivante adéquate : [Λɪ].

La réalisation attendue de la voyelle de CHOICE en anglais standard est [ɔɪ].

La voyelle de MOUTH est une diphtongue qu'on peut symboliser [aʊ] ou [ɑʊ]. Le point de départ le plus postérieur est symptomatique d'un accent raffiné. Nous reviendrons *infra* (Section 6.3.1.2) plus en détail sur les caractéristiques acoustiques de cette voyelle et sur le choix d'un symbole phonétique adapté.

NEAR présente une diphtongue centripète de type [ɪə]. Les réalisations [ɪə:] et [ɪɑ:] traduisent une affection excessive.

Upton ([Upt04], 226) suggère que la forme la plus actuelle de SQUARE est une monophthongue ([ɛ:] ) ou, tout au plus, une monophthongue suivie d'un léger glide vers le centre de l'espace vocalique. La diphtongue [ɛə] stigmatise les formes les plus conservatrices de cet accent. Il y a plus de quarante ans, [Jon66], 64, notait que la réalisation monophthongale n'était qu'occasionnelle.

La voyelle de START est réalisée comme [ɑ:]. Upton ([Upt04], 226) précise que même les locuteurs qui ont une voyelle courte pour BATH, ont une longue dans START. En revanche, ces derniers sont parmi ceux qui ont la réalisation la plus antérieure.

Les ensembles lexicaux NORTH et FORCE se confondent avec THOUGHT en anglais standard. Jusqu'à récemment, les items de FORCE avaient /ɔə/ chez les locuteurs les plus âgés. Voici ce qu'écrivait O'Connor à ce sujet en 1973 ([O'C73], 153) :

/ɔ:/ and /ɔə/ are not often separated : relatively few RP speakers make a contrast, for instance, between 'paw' with /ɔ:/ and 'pour' with /ɔə/ [...] but some still make the contrast and it must be reckoned with.

En ce qui concerne CURE, les formes les plus conservatrices de l'accent ont [ʊə]. Il est de plus en plus fréquent de nos jours d'entendre [ɔ:]. Jones ([Jon66], 66) décrivait un stade intermédiaire entre [ʊə] et [ɔ:], en l'occurrence : [ɔə]. Aux dires de Jones, nombre de locuteurs du sud de l'Angleterre utilisaient déjà [ɔ:] à son époque, surtout dans les mots fréquents. La substitution de [ɔ:] à [ʊə] dans l'ensemble CURE était considérée par Wells ([Wel82], 237 et 287-288) comme un changement en cours ayant pour conséquence, selon le point de vue, une migration d'un grand nombre d'items de CURE vers FORCE-NORTH-THOUGHT, ou même une convergence totale de CURE et du super-ensemble que

nous venons de mentionner, avec toutefois, dans cette seconde perspective, l'existence d'un allophone [ʊə] chez certains locuteurs dans le contexte /Cj\_/. Le dictionnaire de prononciation de Jones ([Jon03]) donne les deux variantes ([ʊə] et [ɔː]) pour la prononciation de <poor>; la monophthongue y est considérée comme la plus usuelle. /ʊə/ est le phonème vocalique le moins fréquent en anglais ([Gim80], 149); ceci constitue probablement un facteur aggravant qui facilite sa disparition, ou la disparition de son allophone principal historique. Gimson en 1980 ([Gim80], 146) ne conseillait que la diphtongue [ʊə] pour les apprenants. En 1998, Nolan ([Nol98], 79) remarquait que lui-même utilisait /ʊə/, et considérait qu'il était peu probable que des mots tels que <sewer> ou <Ruhr> soient prononcés avec /ɔː/.

FIRE et POWER sont des triphthongues que [Upt04], 226, symbolise [ɪɪə] et [aʊə], respectivement. Elles peuvent être réduites à des diphtongues ou à des monophthongues (phénomène du « Smoothing ») chez tous les locuteurs à débit rapide. Curieusement, le Smoothing à débit lent semble traduire une élocution raffinée en RP, alors que c'est également un cockneyisme notoire. Le passage de la triphthongue à la monophthongue peut conduire à une homophonie entre <buyer> (FIRE), <bower> (POWER) et <bar> (START) (exemple tiré de [Gim80], 140).

Pour l'ensemble *happy*, l'anglais standard actuel présente une voyelle tendue, résultant du *happy-tensing*, phénomène détaillé dans la Section 3.2.1.5. On trouve [ɪ] dans une forme conservatrice d'anglais standard.

Les ensembles *letter* et *comma* ont des réalisations en [ə] et appartiennent au même phonème. L'anglais standard n'étant pas rhotique, le /r/ n'apparaît dans *letter* que lorsqu'il sert à faire la liaison avec le mot suivant si ce dernier commence par une voyelle. Traditionnellement, dans un registre très soigné, ce type de liaison était évité au profit d'un coup de glotte : <far away> [fɑː?əweɪ] (exemple repris de [Upt04], 228)<sup>62</sup>. Le r intrusif (une explication plus détaillée est donnée dans la Section 3.2.1.2) semble être la norme, même à l'intérieur d'un mot : <drawing> [dɹɔːɪŋ].

<sup>62</sup>Le symbole utilisé dans la référence correspond en API à une occlusive épiglottale; nous ignorons s'il s'agit d'une coquille, mais avons tout de même noté un coup de glotte.

Cette présentation du système vocalique de l'anglais standard sera confrontée aux données empiriques de notre corpus dans la Section 6.3.1. Avant de clore notre discussion sur cette variété, nous abordons, dans la section suivante, les changements phonético-phonologiques en cours.

### 3.1.2 Changements récents et perspectives

Dans cette section, nous tirons un bilan des évolutions récentes de l'anglais standard et commentons les changements en cours. Les changements récents dont il est question sont ceux qui sont intervenus entre les premières descriptions de Daniel Jones dans les années 1910 et aujourd'hui.

Comme le rappellent très justement Hawkins et Midgley ([HM05]), une étude phonétique avec des sujets classés par tranches d'âge (i.e. transversale) est quelque peu différente d'une véritable étude longitudinale<sup>63</sup>. En effet, la prononciation d'un locuteur donné évolue tout au long de sa vie. L'exemple le plus flagrant (mais peut-être pas le plus représentatif<sup>64</sup>) nous est donné par la reine d'Angleterre. Dans une série de publications, Harrington a examiné les allocutions (annuelles) à l'occasion de Noël de la reine Élisabeth II depuis 1952. Par exemple, dans une étude préliminaire et succincte, Harrington et collègues [HPW00] analysent les deux premiers formants des 11 monophthongues de l'anglais standard produites par la reine dans les années 1950, les années 1980, et par un panel de présentatrices de la BBC enregistrées dans les années 1980. Concernant la prononciation de la reine, l'étude montre qu'au moins un des deux formants varie de façon significative pour 10 des 11 voyelles, et que les deux formants varient pour 5 des 11 voyelles entre 1950 et 1980. La comparaison avec le corpus de présentatrices de la BBC fait apparaître que les valeurs formantiques des enregistrements les plus récents de la reine sont à mi-chemin entre les valeurs de 1950 et celles du corpus de présentatrices. [HPW00] concluent que la

---

<sup>63</sup>Les auteurs anglophones utilisent les adjectifs « apparent-time » et « real-time » pour qualifier les études transversales et longitudinales, respectivement ; voir [Lab94], 43 *sqq.*

<sup>64</sup>Labov commente les résultats de Harrington ([Har06]) et estime pourtant que la reine devrait être particulièrement résistante aux changements phonétiques de la communauté ([Lab06], 501-502).

reine Élisabeth II a inconsciemment modifié sa prononciation pour suivre les tendances inspirées par les locuteurs les plus jeunes d'une forme de prononciation standard du sud de l'Angleterre, moins élitiste que RP. Tout ceci a été rendu possible, entre autres, par l'effondrement d'une stratification sociale rigide, notamment entre les années 1960 et 1980. L'évolution de la prononciation de la reine est particulièrement visible pour la voyelle de TRAP. Guère après la publication de [HPW00], un article paru dans *The Independent* ([Hen00]) dénonçait une étude « disrespectful to the point of *lèse-majesté* ». Le journaliste continuait en remarquant qu'il était bien évident que chacun d'entre nous modifie son accent tout au long de sa vie, et confiait sa vision de l'évolution de la norme de prestige : pour lui, l'accent aristocrate d'il y a cinquante ans a disparu et l'accent des classes moyennes d'hier est devenu la norme de prestige d'aujourd'hui.

Le cas de la voyelle de TRAP présente un intérêt particulier dans une discussion sur l'évolution des systèmes vocaliques. En effet, [Par98], 22, rappelle qu'en 1874, Sweet notait la diffusion d'une réalisation plus fermée dans les classes supérieures. Or depuis quelques décennies, la voyelle de TRAP est de plus en plus ouverte; elle est donc montée puis redescendue dans le triangle vocalique en l'espace d'un siècle environ<sup>65</sup>. Cet exemple appelle plusieurs commentaires. D'abord, si nous nous référons aux principes qui régissent les changements en chaîne (voir Section 2.2.1), on remarque que la fermeture de TRAP est contraire au fait que, dans ce type de changement, on constate généralement que les voyelles brèves s'ouvrent<sup>66</sup>. Ensuite, puisque la « case » laissée vide par TRAP n'a pas été occupée par un autre phonème, et au vu du fait que la pression exercée par TRAP n'a pas été suffisante pour contraindre FLEECE à quitter sa position de voyelle antérieure de degré de fermeture maximale, on peut conclure, avec Labov ([Lab94], 138), qu'il s'agit d'une simple compression des voyelles antérieures. Nous ajouterons qu'il nous apparaît que cette fermeture, puis ouverture, de la voyelle de TRAP en l'espace d'un siècle consti-

<sup>65</sup>On compare parfois la voyelle de TRAP au Grand Old Duke of York (image tirée de [Nol98], 81), dans la comptine éponyme : « The Grand Old Duke of York, He had 10,000 men, He marched them up to the top of the hill, And he marched them down again . . . ».

<sup>66</sup>Les observations empiriques sur lesquelles Labov ([Lab94]) base ces principes sont cependant loin d'être exhaustives. Parallèlement, le statut phonologique bref de la voyelle de TRAP peut être remis en cause pour certains mots d'après [Fud77].

tue certainement un exemple de conflit entre une innovation motivée par des phénomènes sociolinguistiques et les tendances universelles qui semblent régir la « directionnalité » des changements vocaliques, ces dernières ayant fini par l'emporter.

L'antériorisation des voyelles de GOOSE et de FOOT constitue l'une des évolutions les plus marquantes du système vocalique de l'anglais standard ([Wel62, Hen83, HM05, dJMHN07]). Si l'on ajoute à cela l'ouverture plus importante des voyelles antérieures, qui est très marquée pour TRAP, il semble justifié d'émettre l'hypothèse selon laquelle un changement en chaîne, se traduisant par un mouvement des voyelles dans le sens inverse des aiguilles d'une montre, affecte le système des monophthongues de l'anglais standard ([HM05]).

Bien que nous ne traitons pas la question des consonnes, mentionnons brièvement certains traits consonantiques qui pourraient peut-être un jour avoir leur place dans une description de l'anglais standard. Le TH-fronting – la transformation de /θ/ et /ð/ en /f/ et /v/ respectivement – est attesté dans beaucoup de centres urbains en Angleterre (voir [Bri02], 58 *sqq.*). Il en va de même pour le /r/ labiodental ([Nol98, FD00, Mar06]). On sait que ces traits sont perçus comme non-standards<sup>67</sup>, mais on ne peut pas exclure que leur diffusion entraîne une réinterprétation de leur connotation sociale. En effet, ce genre de phénomène est déjà attesté; par exemple, le coup de glotte remplaçant le /t/ intervocalique est stigmatisé dans le sud-est de l'Angleterre comme une forme typique des hommes et de la classe ouvrière. Or il a été démontré, sur la base de l'utilisation croissante de cette variante à Newcastle (remarquable notamment chez les jeunes femmes de classe moyenne), que la réalisation du /t/ intervocalique en coup de glotte correspondait plutôt à une forme de « chic » urbain ([Bri02]).

D'autres innovations concernent le schéma intonatif; par exemple, Nolan ([Nol98], 84) relève chez les locuteurs les plus jeunes l'adoption du « High Rising Terminal » (voir aussi [SH03]). Enfin, le schéma accentuel d'un mot, qui détermine l'alternation voyelles pleines/voyelles réduites, a lui aussi connu quelques changements. Par exemple, d'après

---

<sup>67</sup>À noter néanmoins, concernant le statut du /r/ labiodental, que Nolan ([Nol98], 83-84) l'inclut dans les innovations de RP.

[Nol98], <formidable> est passé de /'fɔ:mɪdəbl/ à /fə'mɪdəbl/. Le mot <armchair> semble également avoir connu un déplacement de l'accent primaire de la seconde vers la première syllabe. Ceci contribue probablement à ce qu'il tende à être interprété comme un tout et non comme deux entités séparables ; de là à imaginer qu'il subisse le même sort que <cupboard> au point où il en viendrait à être prononcé /'ɑ:mtʃə/, cela paraît très improbable étant donné l'omniprésence de l'orthographe dans nos cultures.

## 3.2 Caractéristiques phonético-phonologiques des dialectes des Îles Britanniques

La description géolinguistique qui suit s'appuie en particulier sur les ouvrages de Orton et collègues ([OSW78]), Wells ([Wel82]), Trudgill ([Tru90]), Foulkes et Docherty ([FD99]), et Schneider *et al.* ([SBK<sup>+</sup>04]). L'énumération des traits diagnostiques de chaque variété dans cette section ne présume en rien de leur présence dans nos données ; nous nous contentons de rapporter les propos des spécialistes. Ces éléments sont mis en relation avec les dialectes de notre corpus dans la Partie II. Trudgill ([Tru90]) mentionne le fait que la plupart des Britanniques sont des locuteurs de dialectes modernes, et que si l'on souhaite localiser géographiquement l'un d'entre eux (et c'est précisément notre dessein), il convient d'avoir recours à des critères de prononciation différents de ceux utilisés pour les dialectes traditionnels. Notre analyse des dialectes de l'Angleterre s'appuie sur la classification établie par Trudgill ([Tru90], 50-78). Pour l'intégralité des Îles Britanniques, nous suivrons la typologie de Hughes et collègues ([HTW05]) ; elle est illustrée dans la Table 3.1 et sert de fondement à la description qui suit.

La Table 3.1 reprend la terminologie de Hughes et collègues ([HTW05]) ; or certains des termes employés font référence à des concepts que nous désignerons autrement. En effet, /ʌ/ dans *mud* est un trait qui caractérise les dialectes dans lesquels la scission FOOT-STRUT a eu lieu (voir Section 3.2.1.1). /ɑ:/ dans *path* s'applique aux dialectes qui ont le BATH-broadening, cet allongement accompagné d'une postériorisation de la voyelle

TAB. 3.1 – Classification des dialectes des Îles Britanniques (d’après [HTW05]). Les zones non représentées dans notre corpus sont signalées par \*\*\*.

Aire dialectales	Dialectes de ABI	/ʌ/ dans <i>mud</i>	/ɑ:/ dans <i>path</i>	/ɑ:/ dans <i>palm</i>	/g/ dans <i>sing</i>	/j/ dans <i>few</i>	/i:/ dans <i>hazy</i>	/ɪ/ dans <i>bar</i>	/ʊ/ dans <i>pull</i>	/h/ dans <i>harm</i>	[ei] dans <i>gate</i>
Scot. & North. Ire.	<i>gla/shl/uls</i>	+	-	-	-	+	-	+	-	+	-
S. Ireland	<i>roi</i>	+	+	+	+	+	+	+	-	+	-
Northeast	<i>ncl</i>	-	-	+	+	-	+	+	-	+	-
Central north	***	-	-	+	-	-	+	-	-	+	-
Central Lancashire	<i>lan</i>	-	-	+	-	+	+	-	+	+	-
Merseyside	<i>lvp</i>	-	-	+	+	-	+	-	+	+	+
Humberside	<i>eyk</i>	-	-	+	+	-	+	-	-	+	-
N.W. Midlands	***	-	-	+	-	-	+	-	+	+	+
E. Midlands	***	-	-	+	-	-	+	-	-	-	+
W. Midlands	<i>brm</i>	-	-	+	+	-	+	-	+	+	+
S. Midlands	***	+	+	+	+	-	+	-	-	-	+
E. south-west	***	+	-	-	+	+	+	-	-	+	+
W. south-west	<i>crn</i>	+	-	-	+	+	+	-	-	+	-
South-east	***	+	+	+	+	-	+	-	-	+	+
East Anglia	<i>ean</i>	+	+	+	+	-	+	+	-	-	+
Wales	<i>nwa</i>	+	-	+	+	-	+	-	-	+	-

historique /a/ devant fricative sourde. /ɑ:/ dans *palm* désigne le phénomène qui consiste à avoir une voyelle postérieure longue dans les mots de l’ensemble lexical PALM. /i:/ dans *hazy* fait référence au *happy*-tensing, phénomène décrit plus bas (Section 3.2.1.5). /ɪ/ dans *bar* est plus connu sous le nom de rhoticité. /ʊ/ dans *pull* renvoie à la convergence FOOT-GOOSE, caractéristique des dialectes de l’Écosse et de l’Irlande du Nord, que nous regrouperons sous l’appellation générique « Scots ». Pour /h/ dans *harm*, c’est l’absence de ce trait qui porte un nom : le H-dropping. /g/ dans *sing* dénote la réalisation du digraphe <ng> par [ŋg], différente de la réalisation la plus fréquente : [ŋ]. Concernant /j/ dans *few*, c’est son absence qui est appelée Yod-dropping. Enfin, [ei] dans *gate* fait référence à l’emploi d’une diphtongue fermante pour les mots de l’ensemble lexical FACE. Nous avons inclus dans la Table 3.1 les abréviations des dialectes de notre corpus ; ce dernier est décrit dans la Section 4.1.

L’aspect le plus saillant de cette typologie réside dans le fait que 9 des 10 traits retenus sont d’ordre systémique, le dixième se situant sur le plan de la réalisation. Il semblerait donc que la plupart des caractéristiques essentielles des dialectes des Îles Britanniques

soient d'ordre phonologique et non réalisationnel. Nuancions toutefois : une logique bivalente ne peut guère être utilisée que pour une classification employant comme traits distinctifs des phénomènes phonologiques (car ils sont binaires par nature : présence ou absence du phonème, voyelle arrondie ou non, etc.) ; une valence plus élevée serait nécessaire pour caractériser les aspects phonétiques, et le tableau perdrait en lisibilité.

### 3.2.1 Partition dialectale de l'Angleterre

Notre analyse géolinguistique de l'Angleterre est nettement plus détaillée que celle des autres pays des Îles Britanniques. Il y a deux raisons à cela : il apparaît, d'une part, que la situation est beaucoup mieux documentée, et que les documents sont plus accessibles ; d'autre part, et c'est bien là la raison principale, la majorité des dialectes de notre corpus (huit sur douze, en excluant la variété étiquetée « anglais standard » du calcul) concernent l'Angleterre.

#### 3.2.1.1 Partition nord/sud : scission FOOT-STRUT et BATH-broadening

Nous dirons, en première analyse, que l'Angleterre est traversée par une frontière – dont la localisation précise est malaisée – qui a un statut particulier puisqu'elle est généralement connue de tous les Anglais : « [it] has an acknowledged folk-linguistic status » ([AW04], 178). Cette frontière traduit une réalité qui peut se résumer à l'expression évocatrice « North-South divide » ([Wal00]). Du point de vue de la dialectologie, cette démarcation est matérialisée par deux isoglosses qui, au reste, ne coïncident pas exactement. Le nord est caractérisé par l'absence de scission FOOT-STRUT ; en d'autres termes, la voyelle de ces deux mots appartient au même phonème, dont la réalisation s'approche de [ʊ]. Dans le nord, donc, <luck> et <look> sont homophones. La délabialisation de la voyelle de STRUT (dans certains contextes), qui a conduit à une phonémisation de l'opposition FOOT-STRUT, semble s'être amorcée dans le sud de l'Angleterre vers le milieu du 17<sup>e</sup> siècle ([Wel82], 197). Non seulement cette absence d'opposition FOOT-STRUT est elle-même un trait hautement diagnostique, mais aussi, le phénomène d'hypercorrection qui consiste, pour le locuteur

du nord, à se risquer à une prononciation méridionale en utilisant la voyelle de STRUT pour un mot qui appartient à l'ensemble lexical de FOOT dans le sud, et qui aboutit à des prononciations dont la forme phonologique serait /ʃʌgə/ pour <sugar>, permet de trahir l'origine dialectale de ce locuteur. Knowles ([Kno78], 86), dans une analyse informelle de sa propre prononciation, discerne un second type d'hypercorrection agissant entre les voyelles de deux mots différents, et qui revient à modifier la mauvaise voyelle. Ainsi, dans ce cas, <good luck> et <black castle> en viennent à être prononcés /gʌd lʊk/ et /blʌ:k kæsl/, respectivement.

L'absence de scission FOOT-STRUT est tellement caractéristique du nord que Trudgill ([Tru90], 51) va jusqu'à affirmer qu'elle est connue de « everybody who has spent any time in England » (Soulignement ajouté). La seconde isoglosse qui scinde l'Angleterre en deux est située légèrement au nord de la précédente. Elle sépare le nord linguistique, où la voyelle de BATH est, d'un point de vue systémique, identique à celle de TRAP (avec une qualité proche de [a]), du sud, où a eu lieu le phénomène de BATH-broadening. Ce dernier s'est traduit dans un premier temps par l'allongement de la voyelle de BATH vers la fin du 17<sup>e</sup> siècle, notamment devant une fricative sourde, et a abouti, dans un second temps, à une qualité postérieure de la voyelle au 19<sup>e</sup> siècle. Le nord linguistique, comme le précise Wells ([Wel82], 349) ne correspond pas exactement au nord géographique puisqu'il inclut, en plus de ce dernier, les Midlands.

À l'instar d'un certain nombre d'autres pays, la partition nord/sud dialectologique en Angleterre est également perçue comme une division socio-économique. Il semblerait donc que quand bien même l'on voudrait se concentrer exclusivement sur le facteur géographique, on ne pourrait pas s'affranchir de la dimension sociolinguistique. Par exemple en Italie, le sud véhicule, dans la conscience collective, l'idée de pauvreté et de ruralité. Les traits de prononciation du *mezzogiorno* sont considérés comme diagnostiques de l'appartenance du locuteur à cette région, avec toutes les connotations péjoratives qui en découlent. Le nord, quant à lui, incarne la réussite industrielle et la richesse. En Angleterre, la polarité est inversée : le nord est perçu – et ce cliché est largement entretenu par

les média - comme le royaume des villes crasseuses où règnent les taux de chômage les plus élevés (voir, par exemple, certains films de Ken Loach).

Concernant les deux variables permettant de distinguer le nord du sud, notons d'emblée qu'à l'intérieur du nord linguistique, l'absence de scission FOOT-STRUT et de BATH-broadening ne se manifestent pas de manière homogène. La scission FOOT-STRUT étant tributaire de facteurs sociolinguistiques, Wells ([Wel82], 352) remarque que plus on monte vers le nord, plus le passage d'un système de 5 voyelles brèves à un système à 6 (opposition FOOT-STRUT rétablie dans les classes supérieures) est situé haut dans l'échelle sociale. Ajoutons à cela qu'il existe un stade intermédiaire entre l'absence totale d'opposition FOOT-STRUT et son existence : sur le plan de la réalisation, la forme la plus marquée de l'accent du nord a [ʊ] pour FOOT-STRUT, alors que certains locuteurs tendant vers la variété de prestige (méridionale), mais ne possédant pas le contraste phonologique, vont produire ce qu'on pourrait considérer comme un timbre intermédiaire entre les réalisations de /ʌ/ et /ʊ/ en anglais standard, à savoir, [ə].

Les deux variables marquant la distinction nord-sud n'ont par ailleurs pas exactement le même statut ; Wells ([Wel82], 354) indique :

There are many educated northerners who would not be caught dead doing something so vulgar as to pronounce STRUT words with [ʊ], but who would feel it to be a denial of their identity as northerners to say BATH words with anything other than short [a].

Trudgill ([Tru90], 50-78) propose une classification des dialectes modernes de l'Angleterre qui a l'avantage d'être exposée de façon succincte. Nous suivrons son usage, et conserverons par souci d'homogénéité, sauf pour désigner les quatre points cardinaux, la terminologie anglaise. La Table 3.2 est une matrice de traits binaires offrant une représentation synoptique de la classification opérée par Trudgill ([Tru90]). À partir de 7 traits de prononciation, l'Angleterre peut être divisée en 16 aires dialectales. Ceci n'est d'ailleurs pas tout à fait vrai, puisqu'une inspection sommaire de la Table 3.2 fait apparaître que les traits utilisés ne sont pas suffisants pour distinguer les dialectes étiquetés Northeast et Humberside, et Merseyside et West Midlands, ce qui n'est pas explicitement mentionné

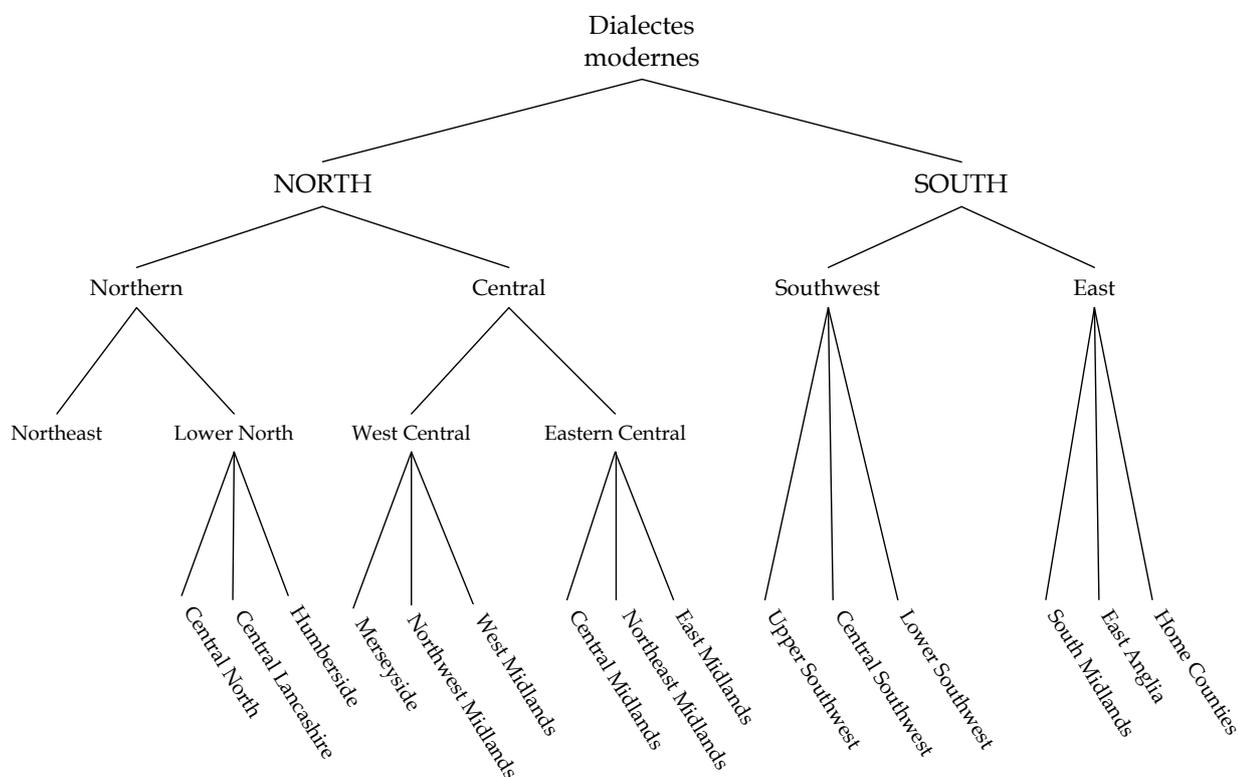


FIG. 3.4 – Regroupement des zones dialectales en supra-régions.

par Trudgill.

Trudgill regroupe ces 16 aires en supra-régions ; nous avons reproduit cette classification arborée dans la Figure 3.4.

Abordons à présent tour à tour chacun des traits employés dans la Table 3.2 : la scission FOOT-STRUT, la rhoticité, la prononciation de [g] dans le digraphe <ng>, le Yod-dropping, le *happy*-tensing, la réalisation de FACE et la vocalisation du /l/. À l'occasion, les paragraphes qui suivent nous donnerons l'opportunité de détailler les caractéristiques

TAB. 3.2 – Classification des dialectes de l'Angleterre (d'après [Tru90]).

Dialectes de ABI	Scission FOOT-STRUT	Rhoticité [g] dans <ng>	Yod-dropping	happy-tensing	monophthongue dans FACE	L Vocalization
Northeast	-	-	-	+	+	-
Central North	-	-	-	-	+	-
Central Lancashire	-	+	-	-	+	-
Humberside	-	-	-	+	+	-
Merseyside	-	+	-	+	-	-
Northwest Midlands	-	+	-	-	-	-
West Midlands	-	+	-	+	-	-
Central Midlands	-	-	-	-	-	-
Northeast Midlands	-	-	-	+	-	-
East Midlands	-	-	+	+	-	-
Upper Southwest	+	+	-	+	-	-
Central Southwest	+	+	-	+	-	+
Lower Southwest	+	-	-	+	+	-
South Midlands	+	-	+	+	-	+
East Anglia	+	-	+	+	-	-
Home Counties	+	-	-	+	-	+

phonétiques d'un ou plusieurs dialectes, notamment lorsque le lien avec les dialectes de notre corpus est évident.

Nous avons déjà traité la question de la scission FOOT-STRUT ; nous ne reviendrons donc pas dessus. Précisons toutefois, comme le fait Trudgill ([Tru90], 66) que le BATH-broadening aurait tout aussi bien pu être utilisé dans cette table en lieu et place de la scission FOOT-STRUT. Voyons le cas de l'accent de Hull, puisqu'il figure dans nos données. Son système vocalique est typique du nord de l'Angleterre car les deux phénomènes que nous venons de mentionner n'ont pas eu lieu. La convergence NURSE-SQUARE est fréquente ([WK99], 146). La voyelle de PRICE a deux variantes : une monophthongue devant consonne voisée (type [a:], apparemment propre à la classe ouvrière [WK99]) et une diphtongue partout ailleurs. Traditionnellement, dans certaines zones du Yorkshire, la voyelle de GOAT est une monophthongue postérieure à mi-chemin entre [ɔ] et [o] ([O'C73], 166-167).

Sur le plan consonantique ([WK99], 147-148), le H-dropping est particulièrement répandu. Le TH-fronting (réalisation de /θ, ð/ en [f, v]) est fréquent dans la jeune génération. Le /r/ labiodental ([ɹ]) existe chez les enfants et les jeunes adultes. /l/ est légèrement vélarisé dans toutes les positions. Le remplacement du [t] (sauf à l'initiale) par un coup de glotte est fréquent chez les locuteurs les plus jeunes.

### 3.2.1.2 La rhoticité

Un accent est dit rhotique lorsque tous les <r> graphiques – et donc souvent historiques – sont prononcés. Ceci s'oppose aux accents non-rhotiques dans lesquels seuls les <r> en position pré-vocalique sont produits. Le <r> pré-consonantique ou en finale de mot a commencé à disparaître dans la prononciation des dialectes du sud-est de l'Angleterre aux alentours du 18<sup>e</sup> siècle<sup>68</sup> ([Tru90], 51, [Wel82], 218 *sqq.*). La perte de la rhoticité

---

<sup>68</sup>Précisons qu'il réapparaît dans les dialectes non-rhotiques en finale de mot lorsque le mot suivant commence par une voyelle ; autrement dit, il sert à l'euphonie qui caractérise les phénomènes de liaison. Ce peut être un <r> orthographique, auquel cas on parle généralement de *linking r*, ou absent de la graphie, et dans ce cas, il s'agit de *intrusive r*. Lorsque le <r> est absent de la graphie et qu'il est néanmoins prononcé, mais pas dans le but de faire la liaison avec un mot commençant par une voyelle, Wells ([Wel82], 343) parle d'*hyper-rhoticité*. Une illustration très connue du r intrusif est *Australia and Asia* : [ɔstreiliə r ən eɪzə]. Autre exemple, apocryphe : *Are you xeroxing it ? No, I'm minoltaing it* ([mɪnɔltæɪŋ]).

continue de gagner du terrain en Angleterre : « Every year the r-pronouncing area gets smaller » ([Tru90], 52). Ceci implique que dans certains dialectes, les locuteurs les plus âgés la conservent, alors que les plus jeunes l'ont perdue. La rhoticité est également marquée socialement. Trudgill ([Tru90], 53) explique que l'introduction précoce de la non-rhoticité dans l'accent de la BBC confère à ce trait un statut social relativement élevé. Ceci revient à dire – et c'est un leitmotiv dans nos travaux – que la dialectologie aréale ne peut pas s'affranchir de considérations ressortissant à la sociolinguistique. En Angleterre, la rhoticité caractérise les dialectes du sud-ouest, ainsi que le Central Lancashire, dans lequel se trouve une des villes où ont eu lieu des enregistrements de notre corpus : Burnley. Si nous nous attardons un instant sur le sud-ouest, nous noterons que les dialectes traditionnels de cette région ont une tendance au voisement des fricatives, d'où l'orthographe humoristique <Zomerzet> ou <Mummerzet>. Par exemple, détail très connu, le féminin de <fox>, en l'occurrence, <vixen>, provient des dialectes du sud-ouest de l'Angleterre, d'où la fricative voisée à l'initiale. Cette prononciation, quoique largement attestée dans le projet VOICES de la BBC<sup>69</sup>, reste l'apanage des locuteurs les plus âgés habitant les endroits les plus reculés ([Elm05], 27). Wells [Wel82] classe la Cornouailles (une partie de notre corpus a été enregistré à Truro) comme faisant partie du sud-ouest. Il s'empresse cependant de préciser qu'elle se singularise par une tradition différente des autres comtés de ce sous-groupe, et par le fait qu'on y parlait le cornique<sup>70</sup>.

La rhoticité, présente donc en Cornouailles, se traduit, sur le plan phonétique, par une approximante rétroflexe, [ɻ]. En plus des traits de prononciation rapportés dans la Table 3.2, on notera que les voyelles de GOAT et FACE semblent être de plus en plus souvent réalisées comme des diphtongues ([Wel82], 347). Wells souligne également que dans certaines parties de la Cornouailles, les voyelles de FOOT et GOOSE ont des qualités antérieures.

---

<sup>69</sup><http://www.bbc.co.uk/voices/>; site consulté le 8 avril 2008.

<sup>70</sup>Les derniers locuteurs natifs du cornique (*kernevek* en langue originale), langue celtique du groupe brittonique, ont vécu à la fin du 18<sup>e</sup> siècle. Une certaine Dolly Pentreath, décédée en 1777 ([Elm05], 5) semble être la toute dernière locutrice native du cornique, et Chesten Marchant, mort un siècle plus tôt, le dernier locuteur monolingue.

### 3.2.1.3 L'occlusive vélaire dans <ng>

La prononciation systématique de [g] dans <ng> est diagnostique des dialectes libellés Central Lancashire, Merseyside, Northwest Midlands et West Midlands. Ces dialectes n'ont pas subi le phénomène que Wells ([Wel82], 188-189) nomme NG-coalescence. Aux alentours de 1600, dans la séquence /ng/ en fin de mot (qui était probablement homorganique : [ŋg]), le [g] a cessé d'être prononcé dans l'accent des classes instruites à Londres. Là où ce phénomène n'a pas eu lieu, la variable [g] dans <ng> est très stable sur toute l'échelle sociale; seuls quelques locuteurs de RP conservateurs ne prononcent pas [ŋg] ([Wel82], 365-366).

Nous allons décrire les variétés de Liverpool et Birmingham, qui appartiennent respectivement aux zones Merseyside et West Midlands. L'accent de Liverpool, le Scouse, se distingue de celui des régions environnantes à cause, dit-on, de l'afflux massif d'immigrés irlandais au 19<sup>e</sup> siècle ([Wel82], 371). On constate, par exemple, qu'en 1851, plus de 22 % de la population de Liverpool étaient composée d'Irlandais ([Hon07]); 25 % en 1861 ([Kno78]). Cette vision est nuancée par Honeybone ([Hon07]) : il note que la plus grande communauté d'immigrants à cette époque reste tout de même celle en provenance de l'Angleterre, et ajoute que d'autres villes, comme Glasgow ou Newcastle, ont elles aussi connu une importante immigration d'Irlandais. Il n'en demeure pas moins que l'accent de Liverpool est ([Kno78], 80) :

an interesting hybrid : on the phonological level, it remains similar to the dialects of neighbouring Northern towns, but phonetically it has been heavily influenced by Anglo-Irish.

En ce qui concerne le système vocalique, le trait le plus typique est la convergence NURSE-SQUARE, plutôt caractéristique de la classe ouvrière. La réalisation de cette unique voyelle est variable : [ɜ:] ou [ɛ:] ([Bea04], 125). La scission FOOT-STRUT n'a pas eu lieu et le BATH-broadening est attesté à partir des classes moyennes ([Wat06], 57).

Sur la plan réalisationnel, il est intéressant de noter que, contrairement à la plupart des dialectes du nord de l'Angleterre, les voyelles de FACE et GOAT sont des diphtongues fermantes; ces diphtongues ont des mouvements formantiques restreints. Les mots terminant en <ook> ont conservé la voyelle longue du moyen-anglais dans une forme dite « very broad » d'anglais liverpuldien, si bien qu'ils appartiennent à GOOSE, et non à FOOT

([Wel82], 373).

Au niveau des consonnes, l'influence de l'anglais d'Irlande se manifeste à travers la réalisation de /θ/ et /ð/ en occlusives dentales ou alvéolaires ([Wel82], 371), phénomène parfois appelé TH-stopping. Paradoxalement, au moins en apparence, les occlusives sourdes (les sonores, plus rarement) en fin de syllabe peuvent être parfois légèrement spirantisées, ou affriquées (et ceci dans toutes les positions).

Passons à présent en revue les traits caractéristiques de l'anglais de Birmingham, souvent appelé « Brummie »<sup>71</sup>. Dans la dichotomie nord-sud, Birmingham est probablement une zone intermédiaire : Wells ([Wel82], 364) mentionne explicitement le fait que FOOT et STRUT puissent être distincts. PRICE et CHOICE peuvent être homophones. La convergence NURSE-SQUARE se rencontre parfois.

Sur le plan réalisationnel, la voyelle de KIT est particulièrement fermée ([O'C73], 155). FACE et GOAT sont des diphtongues à mouvement formantiques amples ([HTW05], 87-88). O'Connor ([O'C73], 154) note que FLEECE a une diphtongue relativement marquée.

#### 3.2.1.4 Le Yod-dropping

Le Yod-dropping caractérise les dialectes East Midlands, South Midlands et East Anglia. La chute du [j] dans les mots <tune>, <dude> et <news>, par exemple, est typique d'un grand nombre de dialectes d'anglais américain ; et le profane considère parfois, à tort, que ce trait distingue ces derniers des dialectes britanniques. En Angleterre, certains contextes consonantiques ont favorisé la disparition de [j] devant [u:] dans une grande majorité de dialectes : lorsque la voyelle était précédée d'une palatale (e.g. <chew>, <juice>), précédée de /r/ (e.g. <crew>, <shrew>) et après une consonne suivie d'un /l/ (e.g. <blue>, <flue>). Ceci correspond à ce que Wells nomme « Early Yod Dropping » ([Wel82], 206). En *General American* s'applique le « Later Yod Dropping » ([Wel82], 247-248), phénomène qui a étendu la chute du [j] devant [u:] aux consonnes coronales. Certains dialectes d'Angleterre, peut-être sous l'influence du prestige que peut dégager la norme

---

<sup>71</sup>Souvenons-nous qu'il s'agit, selon un sondage de la BBC conduit auprès de 5000 personnes, de l'accent que les Anglais affectionnent le moins.

américaine, connaissent une situation proche.

Au-delà, il existe un Yod-dropping total très typique des dialectes mentionnées au début de ce paragraphe, et notamment East Anglia. Cette énumération des traits utilisés dans la taxinomie de Trudgill nous ayant conduit à mentionner l'East Anglia, nous nous arrêtons un instant sur cette variété car elle est présente dans nos données (enregistrements effectués à Lowestoft), et profitons de ce qu'elle a été minutieusement étudiée pour en fournir une description circonstanciée. Le dialecte d'East Anglia est la spécialité du variationniste Peter Trudgill, qui a conduit notamment des études sur la ville de Norwich (située dans le Norfolk) dans les années 1970. L'accent de cette région a joué un rôle capital dans la formation de l'anglais standard. Par ailleurs, si l'on accepte le point de vue selon lequel la langue anglaise est née quand les Anglo-Saxons ont commencé à s'installer en Angleterre, alors on peut considérer que c'est très probablement en East Anglia que l'anglais a été parlé pour la toute première fois dans le monde ([Tru04], 163). Ce dialecte est non rhotique. Entre autres particularités, nous retiendrons (d'après [Tru04]) :

- le Yod-dropping systématique, au point que <dew>/<do> et <Hugh>/<who> sont homophones ;
- l'emploi de /ə/ en position inaccentuée là où on attendrait plutôt /ɪ/ en anglais standard, dans <wanted>, <naked>, etc. Incidemment, ce phénomène est typique de l'accent de Tony Blair (qui n'est pourtant pas originaire de cette région) ;
- l'utilisation, dans les formes les plus archaïsantes, de /ɛ/ pour certains items qui, en anglais standard, appartiennent à TRAP : <catch>, <have>, etc. ;
- certains mots possédant le /ɔ:/ du moyen-anglais peuvent rimer avec *good* : e.g. <road>, <home>, <stone> (/ʊ/), mais pas <coal> (/u:/). L'incidence lexicale de GOOSE et FOOT est différente de celle de l'anglais standard ;
- les diphtongues fermantes – parmi lesquelles Trudgill ([Tru04], 169-172) compte FLEECE et GOOSE – ont la particularité de comporter un second élément plus fermé que dans les autres variétés méridionales ;
- le *happy*-tensing ;

- dans une forme conservatrice du dialecte, les Long Mid Mergers n’ont pas eu lieu, si bien que <days>/<daze> et <maid>/<made> ne constituent pas des paires d’homophones (respectivement pour chaque paire : /æ/ et /e:/);
- pour faire écho au point précédent, et toujours en raison de l’absence de Long Mid Mergers, les voyelles postérieures correspondant au degré d’aperture des deux voyelles que nous venons de mentionner, <nose>/<knows> et <sole>/<soul> ne sont pas homophones (respectivement pour chaque paire : /ʊu/ et /u:/). Trudgill ([Tru04], 170) précise que cette distinction, à la différence de la précédente, est encore bien attestée dans le nord de l’East Anglia;
- la convergence NEAR-SQUARE est attestée dans le nord de l’aire dialectale avec une réalisation proche de [e] ([Tru04], 172);
- certains items de l’ensemble lexical CURE de l’anglais standard ont la voyelle de NURSE. Ainsi <sure> rime avec <her>, et si l’on ajoute à cela le Yod-dropping, <pure> et <purr> sont des homophones dans le nord de la province;
- le Smoothing est largement attesté : <going> est prononcé [gɔ:n], <allow it>, [əlaɪt], etc.

Pour ce qui est des consonnes, la glottalisation des occlusives sourdes en position intervocalique et finale de mot est très répandue; il semblerait d’ailleurs que l’East Anglia soit un des centres à partir desquels ce phénomène s’est généralisé à tout l’anglais britannique. La vocalisation du /l/ est de plus en plus commune dans le sud de l’aire concernée, mais pas au nord. Trudgill ([Tru04], 175) note d’ailleurs que chez les locuteurs du nord de l’East Anglia nés avant 1920, on peut entendre un /l/ clair dans toutes les positions<sup>72</sup>, la distribution sombre *vs* clair actuelle étant la même que celle de l’anglais standard.

### 3.2.1.5 Le *happy-tensing*

Le *happy-tensing* se manifeste de la manière suivante : comme l’explique Wells ([Wel82], 257), la voyelle finale de mots comme <happy>, <coffee>, etc., a été décrite jusqu’aux

---

<sup>72</sup>D’où une très faible probabilité de vocalisation puisqu’il semble que – toute pression sociolinguistique mise à part – seul un /l/ vélarisé puisse aboutir « naturellement » à une voyelle.

années 1950 comme appartenant à KIT. Or sa qualité actuelle est plus proche de [i], ce qui, intuitivement, pousserait à la classer avec FLEECE. D'après la description de Trudgill ([Tru90], 59-60), la version brève de la voyelle est confinée, dans les dialectes modernes, au Central North, Central Lancashire, Northwest Midlands et Central Midlands.

La variante tendue de la voyelle de *happy* soulève le problème du choix du symbole à utiliser dans les dictionnaires de prononciation. En effet, RP avait traditionnellement /ɪ/, et ce n'est vraisemblablement plus le cas aujourd'hui. Les éditeurs du dictionnaire de Jones ([Jon03]), ainsi que Wells dans son dictionnaire ([Wel90]), ont opté pour le symbole /i/. Les premiers considèrent que dans les mots du type de *happy*, il y a neutralisation entre /ɪ/ et /i:/; en disant cela, ils reconnaissent qu'il ne s'agit pas d'un véritable phonème. Si on adopte le principe selon lequel la transcription d'un dictionnaire de prononciation est un équilibre entre un alphabet phonologique (et donc minimaliste) et un guide de prononciation pour les apprenants (cette dernière motivation est au moins implicite à en juger par l'usage de diacritiques), le symbole /i/ peut être justifié. En revanche, dans une optique strictement phonologique, il conviendrait de trancher entre les symboles /ɪ/ et /i:/, et au vu de la proximité phonétique entre la voyelle finale de *happy* et les allophones les plus courants de FLEECE en anglais standard, /i:/ semble plus approprié. Un dernier argument consisterait à dire que la voyelle de *happy* est un phonème à part entière, si on considère que, pour certains locuteurs, <booty> et <bootee> forment une paire minimale<sup>73</sup>.

### 3.2.1.6 La voyelle de FACE

La monophthongue longue dans FACE, d'après la typologie de Trudgill ([Tru90], 60-61), caractérise les dialectes étiquetés Northern et Lower SouthWest. Dans la partie nommée Northeast, le Geordie (accent de Newcastle-upon-Tyne, l'aire dialectale est souvent appelée « Tyneside ») traditionnel et/ou du bas de l'échelle sociale se singularise par la présence d'une diphtongue centripète dans FACE – phonétiquement [iə] – très proche de la

---

<sup>73</sup>Francis Nolan (communication personnelle) a fait mention de la possibilité d'une telle paire.

prononciation de <fierce> en anglais standard, quoique légèrement plus brève. Ceci nous donne l'occasion d'aborder les traits caractéristiques de l'accent de Newcastle. Cette variété n'est pas rhotique. Le /r/ de liaison et le /r/ intrusif sont plus rares que dans d'autres variétés et sont influencés par des facteurs sociolinguistiques (voir [WM99], 31-32). Le /r/ est généralement réalisé comme une approximante alvéolaire, parfois une battue alvéolaire (surtout en position intervocalique). Le « Northumbrian burr », la fricative ou approximante uvulaire, est extrêmement rare et limité aux locuteurs les plus âgés dans les parlers ruraux (elle n'est pas attestée à Newcastle même d'après [Wel82], 368) ; en revanche, le /r/ labiodental ([ʋ]) est de plus en plus répandu, comme dans d'autres variétés britanniques<sup>74</sup>.

Concernant les particularités du système, le Geordie est typiquement septentrional en cela que la scission FOOT-STRUT et le BATH-broadening n'ont pas eu lieu, ce qui revient à dire que nous sommes en présence d'un système à 5 voyelles brèves, sauf, variation sociolinguistique oblige, si un locuteur donné adopte la norme de prestige, en l'occurrence, l'anglais standard. Le dialecte traditionnel avait semble-t-il une convergence NURSE-NORTH, avec NURSE postériorisé jusqu'à [ɔ:] sous l'influence du [ɸ] ([Wel82], 374-375, [Bea04], 126). L'étude instrumentale de Watt et Milroy ([WM99], 38-40) fait apparaître que la réalisation [ɔ:] (donnant potentiellement lieu à une convergence NURSE-NORTH) est utilisée presque exclusivement par les locuteurs les plus âgés de la classe ouvrière (38 % de leurs réalisations), et que ce trait est clairement récessif : les jeunes de la même classe ne l'utilisent que dans 4 % des cas. On note également un potentiel phonème supplémentaire dans des mots tels que <talk>, <walk>, <call>, etc., dont la réalisation est [a:] ([O'C73], 161).

Sur le plan de la réalisation, les voyelles de FACE et GOAT ont un comportement similaire ; il semblerait, en effet, qu'une tendance à la symétrie avant-arrière dans les

---

<sup>74</sup>D'après [FD00], le /r/ labiodental est connu depuis le milieu du 19<sup>e</sup> siècle comme défaut de prononciation chez les enfants. Gimson, en 1980, y fait encore référence en qualifiant ce trait de « defective » ([FD00], 31). Paradoxalement, ce trait de prononciation semble également avoir été naguère symptomatique d'un style affecté typique des classes supérieures. La variante a désormais sa place dans les médias, puisque on peut voir dans une publicité pour de la lessive, un présentateur lancer son fameux « all right » avec un /r/ labiodental, jouant ainsi sur l'ambiguïté possible avec « all white ». Même Margaret Thatcher aurait adopté consciemment cette variante afin d'adoucir son image ([FD00], 33). L'innovation semble provenir du sud-est de l'Angleterre, et elle s'est ensuite diffusée au reste de l'Angleterre, mais pas encore, semble-t-il, aux autres pays des Îles Britanniques.

systèmes vocaliques soit universelle ([WM99], 32). Les variantes les plus attendues, et donc non marquées, sont de longues monophthongues du type [e:] et [o:]. Dans la classe ouvrière, en particulier chez les locuteurs masculins les plus âgés, on trouve les réalisations les plus conservatrices (en cela qu'elles correspondent au dialecte traditionnel de la zone concernée), i.e. les diphtongues centripètes [ɪə] et [ʊə] ([WM99], 34-38). La voyelle de NURSE possède trois variantes dans les descriptions traditionnelles :

- une voyelle postérieure de type [ɔ:] qui pourrait, dans certains cas, traduire une convergence partielle ou totale avec les ensembles THOUGHT, NORTH et FORCE. Watt et Milroy ([WM99], 33) regardent cette éventualité avec circonspection. Cette variante est l'apanage des locuteurs masculins les plus âgés de la classe ouvrière ;
- une variante proche de l'anglais standard, [ɜ:] ;
- une réalisation antérieure, arrondie à des degrés variables, [ø:], variante de prédilection des femmes jeunes, quelle que soit leur classe sociale.

La réalisation de la voyelle de PRICE est, comme à Hull (Section 3.2.1.1), conditionnée par l'environnement consonantique. Le premier élément de la diphtongue est plus fermé devant une occlusive ou une fricative sourde. Par exemple, <knife> est réalisé [nɛɪf], et <knives>, [naɪvz] ([WM99], 28-29).

### 3.2.1.7 La vocalisation du /l/ (L-vocalization)

Chez Trudgill ([Tru90], 60-62), la vocalisation du /l/ est diagnostique des Home Counties. Ce trait est typique d'un accent connu sous le nom de Estuary English<sup>75</sup>. Le /l/ en position pré-consonantique est vélarisé dans nombre de variétés d'anglais : au contact de l'apex avec les alvéoles s'ajoute un renflement du dos de la langue en direction du voile du palais, ce qui confère à [ɫ] une qualité proche de [o] ou [u]. Wells ([Wel91]) reconnaissait à demi-mot que ce phénomène pouvait désormais être inclus dans les descriptions de l'anglais standard. On se souviendra que la vocalisation du /l/ s'est produite de ce côté-ci

---

<sup>75</sup>C'est un terme qui désigne depuis une trentaine d'années l'accent caractéristique de l'estuaire de la Tamise qui, depuis, s'est diffusé au-delà.

de la Manche il y a plus de mille ans<sup>76</sup>.

### 3.2.2 Le Pays de Galles

Nous commencerons cet exposé par un bref aperçu de la situation, non pas de l'anglais du Pays de Galles, mais du gallois. Le gallois est le descendant direct de la langue celtique parlée sur la plus grande partie du sol britannique avant les invasions anglo-saxonnes. Il comptait 21 583 locuteurs monolingues en 1981, soit environ 0,8 % de la population totale du Pays de Galles ([DO97], 584). 90 ans plus tôt, cette fraction atteignait environ les 30 %. Fennell ([Fen01], 197) constate qu'il ne reste plus un seul locuteur de gallois monolingue aujourd'hui<sup>77</sup>. En ce qui concerne les locuteurs bilingues anglais/gallois, ils représentaient 50 % de la population du pays au début du 20<sup>e</sup> siècle et moins de 20 % en 1981 ([DO97], 585). Toutefois, comme c'est le cas pour les autres langues celtiques encore parlées sur les Îles Britanniques, le gallois connaît un renouveau grâce à l'appui des pouvoirs publics et des médias. Sans entrer dans une analyse profonde de ce phénomène, nous livrerons les résultats éloquentes d'une étude de Drinkwater et O'Leary ([DO97], *passim*). À partir d'un examen du *Sample of Anonymised Records* extrait du *1991 Population Census*, les auteurs ont constaté, méthodes quantitatives à l'appui, que les locuteurs du gallois au Pays de Galles sont moins touchés par le chômage que leurs compatriotes qui ne possèdent aucune compétence dans cet idiome. Il se pourrait que ceci soit partiellement imputable à un phénomène de discrimination positive à l'embauche procédant du renforcement d'un sentiment patriotique. Cependant, Penhallurick ([Pen04], 99) soutient que, malgré les efforts récents, le gallois tend vers une « latinisation », processus par lequel une langue vient à n'être utilisée que dans un nombre très restreint de domaines.

En première approximation, notons que l'anglais du Pays de Galles peut être divisé en

---

<sup>76</sup>Il semblerait que le /l/ latin devant consonne fût vélarisé. La perte du contact apical s'est amorcée, d'après Joly ([Jol95],100) aux alentours du 7<sup>e</sup> siècle. [ɫ] a semble-t-il terminé de se vocaliser en [u] au 10<sup>e</sup> siècle, formant ainsi une diphtongue avec la voyelle qui le précédait.

<sup>77</sup>Il semblerait néanmoins que le gallois existe encore en tant que langue maternelle : Mark Jones (communication personnelle) rapporte que, alors qu'il effectuait un travail de terrain au Pays de Galles, il a constaté que certains jeunes privilégiaient le gallois jusqu'à ce qu'ils entrent dans la vie active.

deux variétés, l'une typique du nord-ouest, l'autre, du sud. Ces deux variétés sont, d'après Penhallurick ([Pen04], 99), le reflet des deux dialectes du gallois. Formulé autrement, cela revient à dire que le substrat gallois a constitué un facteur important dans la partition dialectale de l'anglais au Pays de Galles. Les variétés parlées en Angleterre dans les zones frontalières ont également influencé l'anglais du Pays de Galles. La description de Penhallurick ([Pen04]) s'appuie sur une variété étiquetée « traditional, rural Welsh English ». Nous ne saurions donc être trop prudent quant à la possibilité de généraliser ce qui suit. L'anglais du Pays de Galles n'est pas rhotique, mais l'influence du gallois dans les zones où ce dernier est encore très présent (i.e. la moitié ouest), la proximité géographique avec des accents d'Angleterre rhotiques (i.e. au sud-ouest) font que des /r/ peuvent être réalisés par certains locuteurs dans toutes les positions. Voici quelques traits caractéristiques :

- certains mots appartenant à LOT en anglais standard, mais orthographiés avec un <a> (<quarry>, <wash>, <wasp>, etc.), ont la voyelle de TRAP ;
- Wells ([Wel82], 380) mentionne la convergence STRUT-Schwa comme caractéristique incontournable. Penhallurick ([Pen04], 103) nuance le propos en parlant de « the blurring or even erasing of distinction » (Soulignement ajouté) ;
- au nord-est, certains mots de l'ensemble STRUT de l'anglais standard ont la voyelle de FOOT, typique des accents du nord de l'Angleterre ; cela semble dû à la proximité avec le Cheshire ;
- certains mots de l'ensemble ONE (cet ensemble lexical peut être considéré comme un sous-ensemble de STRUT regroupant les voyelles orthographiées <o>) sont prononcés avec la voyelle de LOT ; c'est très vraisemblablement le résultat d'une prononciation orthographique ;
- la voyelle de BATH est généralement réalisée en [a], mais il existe une variante allongée [a:], voire allongée et postérieure [ɑ:]. La réalisation précise semble être corrélée au registre de langue et à la classe sociale ([Pen04], 104) ;
- la situation de FACE-STAY et GOAT-SNOW est complexe car elle est loin d'être homo-

gène sur tout le territoire. Sans entrer dans les détails<sup>78</sup>, retenons que le gallois n'a pas de diphtongue de type /ei/ et /ou/. Par conséquent, la phonétique du substrat influençant ici celle du superstrat, les voyelles de FACE et GOAT sont réalisées comme de longues monophthongues. Ce phénomène est, une fois de plus, renforcé au nord du Pays de Galles par la proximité avec les accents de l'Angleterre septentrionale, qui ont en général également une monophthongue. L'orthographe influence les deux voyelles qui nous intéressent dans l'autre direction : les graphies <ai>, <ay>, <ei> et <ey> ont tendance à engendrer la prononciation [ei], et <ou> et <ow> tendent à aboutir à [ou], d'où la nécessité d'identifier deux ensembles lexicaux supplémentaires : STAY et SNOW ;

- les diphtongues de PRICE, CHOICE et MOUTH<sup>79</sup> ont un second élément particulièrement fermé. Il semble y avoir deux réalisations de ces voyelles : [ai] et [əi] pour PRICE, et [au] et [əu] pour MOUTH. Les variantes avec un schwa semblent confinées au sud et sud-est, ce qui peut laisser penser, comme le note Penhallurick ([Pen04]) que les deux variantes de chaque ensemble sont déterminées par l'époque à laquelle l'anglais s'est implanté dans l'aire concernée ; le sud-est étant la première zone à avoir subi l'influence de l'anglais. Ceci est tout à fait en accord avec la chronologie du Grand Changement Vocalique ;
- CURE, POWER et FIRE sont très clairement dissyllabiques, avec un glide marqué intervenant entre les deux éléments de la diphtongue. Ce trait est hautement caractéristique, et c'est en cela que l'anglais gallois s'oppose à d'autres dialectes, comme East Anglia, dans lesquels se produit le phénomène de Smoothing ;
- l'anglais du Pays de Galles a tendance à avoir, en termes phonétiques, plus de monophthongues et moins de diphtongues que la plupart des autres dialectes ([Wel82], 379) ; par exemple, contrairement à un certain nombre d'autres dialectes, FLEECE

---

<sup>78</sup>Nous reprenons l'argumentaire de [Pen04], 105-106, ainsi que les symboles qu'il utilise.

<sup>79</sup>[Pen04], 106, utilise MOUSE et non MOUTH (étiquette conventionnelle) pour désigner cet ensemble lexical. Or ces deux mots contiennent visiblement, d'après l'*OED*, la même voyelle historique et le contexte consonantique est tellement similaire qu'il est fort peu probable que l'auteur ait souhaité désigner autre chose que MOUTH. Étant donné qu'à la page 101 dans [Pen04], c'est bien MOUTH qui est employé, nous avons rétabli ici l'appellation conventionnelle de cet ensemble lexical.

et GOOSE sont des monophthongues ;

- la voyelle de NEAR peut être prononcée [jɜ:] d’après [Wel82], 380, auquel cas <fear> rime avec <fur>.

Sur le plan consonantique, notons que les occlusives sourdes sont fortement aspirées, surtout au nord. La plupart des consonnes sont allongées, sous l’influence du gallois, en position médiane de mot ; ceci est également particulièrement singulier. L’articulation dentale (plutôt qu’alvéolaire comme en anglais standard) pour les apicales /t, d, n/ semble être la norme au nord. Les fricatives initiales peuvent être voisées comme c’est le cas dans certains accents du sud-ouest de l’Angleterre. /r/ est généralement réalisé en approximante (comme en anglais standard). Les roulées et les battues ne sont cependant pas rares. On note aussi sporadiquement des réalisations uvulaires dans le nord. Wells rapporte que le [l] clair est très répandu, surtout au sud ([Wel82], 379).

### 3.2.3 L’Écosse

D’après Wells ([Wel82], 394), l’anglais écossais est un anglais d’Angleterre standard prononcé avec un accent écossais, qui coexiste avec le scots, un peu comme un dialecte moderne peut coexister avec un dialecte traditionnel dans le nord de l’Angleterre. Stuart-Smith ([SS04], 47) parle de continuum corrélé à la classe sociale allant de « broad Scots » à « Scottish Standard English » (ScStE), le scots tendant à être restreint à certaines situations de communication informelle.

Il est généralement admis que l’anglais d’Écosse n’a pas de durée vocalique distinctive dans son système ; c’est d’ailleurs une caractéristique qui sera illustrée dans notre étude sur le rythme (Chapitre 5). En revanche, certains contextes consonantiques ou morphémiques ont une influence particulière sur la durée des voyelles ; c’est la *Scottish Vowel Length Rule* (SVLR), également connue sous le nom de loi d’Aitken (voir par exemple [McC77, SHT99]). Lorsqu’une voyelle précède une fricative voisée, un /r/, ou une frontière morphémique, elle est phonétiquement longue. D’après Scobbie *et al.* ([SHT99]), la SVLR ne s’applique qu’aux voyelles /i/, /u/ et /ai/. Par exemple, les voyelles de *breathe*,

*beer*, *bee* et *agreed* sont plus longues que celles de *brief*, *bead* et *greed* ([SS04], 57).

Pour rester à un niveau systémique, notons encore que les voyelles de FOOT et GOOSE ne sont pas distinctes ; ce phénomène existe également en Ulster et dans l'extrême nord du Northumberland ([Wel82], 402). Le BATH-broadening n'a pas eu lieu. La plupart des locuteurs de ScStE ont un seul phonème pour la voyelle de LOT, CLOTH et THOUGHT, faisant de <cot> et <caught> des homophones. Incidemment, cette convergence caractérise les variétés de l'ouest des États-Unis ([Gor04], 342). La convergence NORTH-FORCE n'a pas eu lieu pour l'échantillon de ScStE de Glasgow décrit dans [SS04], 53. Il existerait également, de façon sporadique, un phonème intermédiaire entre KIT et DRESS, auquel [SS04], 57 fait référence avec le mot-clé NEVER.

La voyelle de FACE et son homologue postérieur, GOAT, sont généralement réalisés comme des monophthongues. La voyelle de OUT semble être un indicateur assez fiable de la classe sociale : l'usage d'une monophthongue de type [ɥ] est typique de la classe ouvrière, une diphtongue étant de plus en plus fréquente au fur et à mesure qu'on remonte l'échelle sociale. La voyelle de *happy* n'est non seulement pas tendue, mais sa réalisation est plus ouverte, proche de [e].

Les variétés d'anglais d'Écosse sont généralement rhotiques ; Stuart-Smith ([SS04], 62) rapporte qu'environ 90 % des /r/ post-vocaliques de l'une de ses études sont articulés chez les locuteurs de classe moyenne<sup>80</sup>. Les réalisations de /r/ sont très variables : une approximante post-alvéolaire (comme en anglais standard), une approximante rétroflexe, une battue, ou, plus rarement, une roulée. Il arrive parfois que /r/ soit réalisé en [ə]. Wells ([Wel82], 411) note la possibilité d'une fricative uvulaire comme variante idiosyncrasique. Le choix de ces variantes est conditionné par la position à l'intérieur du mot, le contexte phonétique, le sexe, ou encore la classe sociale (pour plus de détails, se référer à [SS04], 62-63, [SS99], 210, [Chi99], 228, [Wel82], 410-411).

En ce qui concerne les consonnes, on retiendra que les apicales peuvent être soit alvéolaires, soit dentales. ScStE possède deux phonèmes consonantiques en plus de la plupart

---

<sup>80</sup>Une étude récente de Stuart-Smith ([SS07]) montre une dérhoticisation chez les adolescents à Glasgow.

des variétés d'anglais, /ʌ/ et /x/ (pour une discussion détaillée, voir [Chi99], 227 et [SSS06]). Le /l/ est généralement vélarisé dans toutes les positions du mot ; mais un [l] clair est possible, notamment dans les Hautes-Terres, avec une distribution clair *vs* sombre identique à l'anglais standard.

Pour ce qui concerne la dialectologie aréale, Wells ([Wel82], 397) propose 3 zones :

- les dialectes centraux des Basses Terres, incluant principalement la région d'Édimbourg (Lothians) et celle de Glasgow (Strathclyde) ;
- les dialectes méridionaux, avec la région de Dumfries-and-Galloway à l'ouest, et les Borders à l'est ;
- les dialectes septentrionaux, au nord de Stonehaven.

Stuart-Smith ([SS04], 47), quant à elle, remarque que le ScStE (anglais standard parlé avec un accent écossais) ne varie que dans une très faible mesure sur tout le territoire, alors qu'on peut distinguer quatre variétés de scots : « Mid or Central Scots, Southern or Border Scots, Northern Scots, and Insular Scots ».

Il nous est particulièrement malaisé à ce stade de notre étude d'entrevoir quels pourraient être les traits (notamment vocaliques) qui permettraient de distinguer les deux dialectes dont nous disposons dans notre base de données : celui des Hautes-Terres et celui de Glasgow.

### 3.2.4 L'Irlande

Bien que nous traitions les deux Irlandes dans la même section, souvenons-nous qu'il s'agit de deux zones dialectales très différentes.

Wells ([Wel82], 417) rappelle les 3 principales sources d'influence de l'anglais irlandais actuel :

- l'anglais venu de l'ouest de l'Angleterre ;
- le scots et l'accent qui lui est attaché ;
- le erse (gaélique irlandais), langue indigène, dont l'influence en tant que substrat semble incontestable.

Wells ([Wel82], 418) insiste sur le fait que l'anglais irlandais est demeuré relativement conservateur en cela qu'il est resté imperméable à nombre d'innovations tant britanniques qu'américaines. Nous calquons la partition dialectale sur la division politique et démographique, qui oppose l'Ulster (ou Irlande du Nord) à la République d'Irlande. Ces régions sont représentées dans notre corpus par les villes de Belfast et Dublin, respectivement. Certes ces supra-régions dialectales peuvent être subdivisées en plusieurs zones, mais une granularité plus fine dépasserait les limites imposées par notre corpus.

À l'indépendance politique de la République d'Irlande (vis-à-vis du Royaume-Uni) semble correspondre une indépendance (socio-) linguistique. Ceci se manifeste à travers le fait que, pour un Irlandais, l'anglais standard du sud de l'Angleterre n'est pas vu comme une norme de prestige, mais plutôt comme une variété que Hickey ([Hic99], 265) qualifie de « extra-national » et de « definitely un-Irish »<sup>81</sup>. Ce rejet de l'anglais standard n'a, par exemple, pas été aussi ferme en Écosse.

À des fins typologiques, il peut parfois être intéressant de regarder l'Irlande comme une seule entité linguistique. Dans ce cas, Hickey, dans *A Sound Atlas of Irish English* ([Hic04a]), établit une liste de traits de prononciation partagés par la plupart des variétés vernaculaires le l'île :

- la lénition des occlusives alvéolaires, e.g. <city> [sɪrɪ];
- le [l] clair dans toutes les positions (ceci se limite aux variétés les plus conservatrices);
- la rhoticité : dans le sud, la réalisation traditionnelle est une approximante alvéolaire vélarisée. Au nord, on rencontre une approximante rétroflexe, comme en Écosse;
- la distinction entre les voyelles brèves devant /r/ dans les variétés conservatrices : <term> [tɛɪm] vs <turn> [tʌɪm];
- la distinction entre /w/ et /ʌ/ (qui oppose, par exemple, <witch> et <which>);
- l'absence fréquente de distinction entre les voyelles de TRAP et BATH.

Si, comme le rappelle Hickey ([Hic04a], 41), certains de ces traits ont un potentiel

---

<sup>81</sup>Lors d'une conférence, nous avons, afin d'être concis, remplacé la formulation « dialectes des Îles Britanniques » par « dialectes britanniques ». Une participante nous avait vivement reproché ce raccourci qu'elle jugeait maladroit.

typologique limitée (puisqu'ils existent également dans d'autres variétés), la non rhoticité ou encore le *h*-dropping permettent de déduire sans hésitation que le locuteur n'est pas irlandais. Comme nous l'avons remarqué plus haut, l'Irlande du Nord et la République d'Irlande sont deux régions linguistiques bien distinctes ; nous conserverons donc la dichotomie dans notre exposé. Nous garderons également l'opposition entre variétés urbaines et rurales.

### 3.2.4.1 Le sud

Sur le plan systémique, Wells ([Wel82], 421) notait un phénomène particulièrement intéressant à rapprocher des ensembles lexicaux *NURSE* et *SQUARE* de l'anglais standard. Les mots-clés <pair>, <per> et <purr> possèdent trois, deux phonèmes différents ou un seul dans les variétés étiquetées « typical southern Irish provincial », « typical Dublin » et « smart Dublin », respectivement. L'auteur constate que la convergence *NURSE-SQUARE*, trait stigmatisant à Liverpool, est, au contraire, perçue favorablement par certains Dublinois. Hickey ([Hic04b], 91) considère la convergence *NURSE-SQUARE* comme possible dans ce qu'il intitule « fashionable Dublin » et habituel en « rural northern ». *FORCE* et *NORTH* sont distincts dans les 5 variétés présentées par Hickey ([Hic04b], 91), sauf potentiellement en « fashionable Dublin ».

À Dublin, traditionnellement, les voyelles longues fermées en syllabe fermée sont réalisées comme deux syllabes séparées par un hiatus : <clean> [kliːən], <fool> [fuːwəl] ([Hic04a], 45). Le /ʊ/ antérieur à la scission *FOOT-STRUT* est conservé, e.g <Dublin> [dʊblən]. La voyelle de *LOT* est allongée lorsqu'elle est suivie d'une fricative sourde ; c'est l'ensemble *CLOTH* (Hickey l'appelle *SOFT*) qui se distingue de *LOT* ici<sup>82</sup>.

Hickey ([Hic04a], 46-50) note la formation d'un nouveau dialecte urbain à Dublin ces vingt dernières années. Le grand nombre d'immigrants, dû à une prospérité sans précédent de la ville sur cette période, constitue un groupe d'individus socialement mobiles qui cherchent à se différencier de ce qu'ils perçoivent comme le dialecte populaire. Ceci a eu

---

<sup>82</sup>Souvenons-nous que les mots de l'ensemble *CLOTH* avaient naguère une voyelle longue et relativement fermée dans une forme très conservatrice de RP (voir Section 3.1.1).

TAB. 3.3 – Changement en chaîne récent à Dublin. D’après ([Hic04a], 47)

<cot>	[kɒt] → [kɔt]
<caught>	[kɒ:t] → [kɔ:t]
<time>	[tʌɪm] → [tʌmm]
<toy>	[tɔɪ] → [tɔɪ], [toɪ]

pour conséquence un véritable changement en chaîne dans le système vocalique, qui (avec d’autres traits de prononciation) tend à se diffuser au-delà de la métropole puisque c’est cette forme de prononciation que vont adopter de fait les locuteurs de l’Irlande du Sud à la recherche d’un accent non régional. Ce changement en chaîne se traduit par un mouvement des timbres dans le sens inverse des aiguilles d’une montre, ce mouvement étant également répercuté sur le premier élément des diphtongues. Les effets de ce changement sont récapitulés dans la Table 3.3.

Ce changement est encore plus marqué lorsque la voyelle précède /r/ ; à Dublin, c’est une approximante rétroflexe : [ɻ]. Cette nouvelle forme de prononciation possède en outre, parmi les traits vocaliques, un premier élément antérieur et relativement fermé dans MOUTH : [æʊ] ou [ɛʊ].

Le TH-stopping (également attesté à Liverpool) est cantonné à l’Irlande du Sud ([Hic04a], 38). Il est attesté dans le dialecte urbain traditionnel de Dublin au moins depuis le 17<sup>e</sup> siècle ([Hic04a], 45). La lénition du /t/ en position intervocalique après une voyelle accentuée est également typique de cet accent. Des réalisations diverses montrent que ce processus peut atteindre différents niveaux de complétude selon le locuteur : <water> [wɑ:t̪ə] → [wɑ:ɹə] → [wɑ:hə] → [wɑ:ə]. Le /l/, traditionnellement clair dans toutes les positions, semble avoir une variante sombre dont la distribution est celle de l’anglais standard ([Hic04a], 49).

### 3.2.4.2 Le nord

Sur le plan du système vocalique, on remarque une possible homophonie entre <fir>, <fair> et <fur>. Si l'on s'en tient à la comparaison avec l'anglais standard, cela implique une convergence NURSE-SQUARE ; <fir> et <fur>, quant à eux, appartiennent déjà à l'ensemble NURSE en anglais standard. Contrairement à l'Irlande du Sud, on trouve au nord la convergence FOOT-GOOSE, caractéristique de l'anglais d'Écosse ([Wel82], 437). Milroy et Milroy ([MM78], 25) distinguent un ensemble lexical composé de 22 items (<pull>, <took>, <look>, etc.) comportant une voyelle tantôt réalisée [ʌ], tantôt [ɯ]. La durée vocalique semble ne pas être pertinente dans le système, mais plutôt influencée, comme c'est le cas en Écosse, par le contexte phonétique (pour plus de détails, voir [Wel82], 439-440). Les ensembles TRAP, BATH et PALM ont le même phonème. Les ensembles LOT et THOUGHT ont parfois deux phonèmes différents.

Au niveau de la réalisation phonétique, la voyelle de DRESS tend à être produite comme une diphtongue centripète : [eə]. La voyelle de *happy* s'approche de [e]. La voyelle de l'ensemble FOOT-GOOSE présente un timbre antérieur.

Concernant les consonnes, le TH-stopping est restreint aux cas où /θ/ et /ð/ sont suivis de /r/ (cf. 3.2.4.1, 3.2.1.3).

À Belfast, d'après Hickey ([Hic04a], 51), on a sur le plan de la réalisation :

- le son [ɑ] pour LOT ;
- la voyelle [ɛ:] dans <beat> ;
- une ouverture de DRESS ;
- /aʊ/ avant /l/ dans les monosyllabes, e.g. <old> [aʊl] ;
- TRAP se ferme devant les vélares (<back> [bɛk], <bag> [bɛg]) et après /k/ (e.g. cap [kɛp]) ;
- FLEECE et GOOSE ont des réalisations brèves ;
- KIT est plus ouvert, voire centralisé.

Notre revue des particularités phonético-phonologiques des dialectes des Îles Britanniques est loin d'être exhaustive. La multiplicité des facteurs de variation – âge, profil

sociologique, dialecte urbain *vs* rural – nous a conduit à simplifier une réalité hautement complexe. Nous nous sommes cependant efforcé de lister les principaux phénomènes phonologiques et réalisationnels susceptibles d'intervenir dans l'interprétation de l'analyse acoustique de la Partie II.