



HAL
open science

Interfaces and representations in English phonology

Quentin Dabouis

► **To cite this version:**

Quentin Dabouis. Interfaces and representations in English phonology. Linguistics. Université de Tours, 2024. tel-04827041

HAL Id: tel-04827041

<https://hal.science/tel-04827041v1>

Submitted on 9 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HABILITATION À DIRIGER DES RECHERCHES

**INTERFACES AND REPRESENTATIONS
IN ENGLISH PHONOLOGY**

Volume I : Document de synthèse

Sciences du Langage

Année universitaire : 2024 / 2025

présentée et soutenue publiquement par :

Quentin Dabouis

le 6 décembre 2024

JURY :

(Par ordre alphabétique)

Prénom	NOM	Grade*	Établissement d'exercice
M. Ricardo	BERMÚDEZ-OTERO	Senior lecturer	University of Manchester
Mme Sandrine	FERRÉ	PR	Université de Tours
M. Jean-Michel	FOURNIER	PR	Université de Tours
M. Olivier	GLAIN	PR	Université Jean Monnet Saint-Etienne
M. Ingo	PLAG	Professor	Heinrich-Heine-Universität Düsseldorf
Mme Anne	PRZEWOZNY-DESRIAUX	PR	Université Toulouse Jean-Jaurès

Acknowledgements

First, I would like to warmly thank Jean-Michel Fournier, who brought me into the world of English phonology and has agreed to accompany me one last time on this project. I am also very grateful to the members of the jury, Ricardo Bermúdez-Otero, Sandrine Ferré, Olivier Glain, Ingo Plag and Anne Przewozny-Desriaux, for accepting to evaluate this work. I am very honoured by and thankful for their presence on this jury.

The collective aspect of scientific research is crucial to me, not only peer-reviewing and conferences, but collaborations of all kinds, exchanges of ideas and informal conversations. Many of the studies that are reported in this document owe a lot to the contributions of colleagues from all over the world, be it through direct collaboration or a suggestion after a conference presentation, and so I am thankful to all who have contributed to making those studies what they are. I have had the chance of collaborating with a number of different colleagues over the years, and so I am grateful to all my co-authors for the work that we did together: Jean-Michel Fournier, Pierre Fournier, Véronique Abasq, Marjolaine Martin, Sabine Arndt-Lappe, Nicola Lampitelli, Guillaume Enguehard, Olivier Glain, Sylvain Navarro, Marie Gabillet, Aaron Seiler and Isabelle Girard. It has been a pleasure to work with each and every one of you! Four colleagues have kindly accepted to read small sections and have provided useful feedback: Nicola Lampitelli on Strict CV, Péter Szigetvári on the vowel system of SSBE and stress and accent and Renate Raffelsiefen and Heather Newell on suffix classes. All errors are mine alone.

In the past two years, one group of collaborators has been central to my research activities: the members of the ERSaF project. It has been great to work with Sabine, Marie and Aaron, and I look forward to the last year of the project. I am also thankful to the colleagues that we have met for taking the time to meet with us to discuss our project, for their ideas and suggestions: Claire Moore-Cantwell, Ricardo Bermúdez-Otero, Ingo Plag, Birgit Alber and Rebecca Treiman.

Ever since I started working as a researcher, I have been exchanging with a few colleagues who kindly took the time to answer my numerous emails. Most prominent of them is Ricardo Bermúdez-Otero, who has been answering my tricky questions about foot structure and Stratal Phonology for close to a decade. Thanks again for those long emails and fascinating discussions! Thanks to Ingo Plag and Renate Raffelsiefen, who answered several long emails on English morphology, to Christophe dos Santos, who introduced me to Prosodic Phonology and Optimality Theory and answered many of my questions when I was a PhD student. Finally, thanks to Geoff Lindsey for our recent exchanges on the vowel system of SSBE.

The material conditions in which I have been able to work this year in order to be able to finish this document in time were also crucial, and so I am very thankful to my colleagues of the English department of the Université Clermont Auvergne, who took over some of the responsibilities that I had been in charge of and made it possible for my teaching load to be compatible with an increased time spent on research. I am also thankful to my colleagues from the Laboratoire de Recherche sur le Langage for their support, and especially Paul Lotin for his help with the extraction of audio data from YouGlish. I am also thankful to Axel Delarue for his assistance with statistical analyses.

Last but not least, I am very thankful to my close ones, who have supported me over the years. Most prominent of all is Anaïs, with whom I have had the chance of sharing the last three years, and who made them unlike any other! Your continuous support has been invaluable. Everyday life has also been made more beautiful by sharing a roof with many great people: Kristell, Etienne, Lucas, Gautier, Amandine, Manon, Marie-Camille, Brigitte and our furry friends Salto, Kopine, Romy, Patoune and Doucette! Music has also been a space to let some steam off and evolve in an ecosystem that is very different from that of scholarly life. Thanks to Florent, Anthony, Clément, Quentin, Guillaume, Jean-Baptiste, Lucas, Nicolas, Thomas, Paul, Julien and Pierre-François for those good times! Life in Clermont-Ferrand has been made sweeter and more interesting by a number of people, whose list would be too long, but here are a few: Gaïa, Damien, Antoine, Quentin, Annelise, Jeanne, Sofiane, Adeline, Coralie, Luca, Lucille, Samuel, Anne, Marion... And to conclude this overlong list of thanks, I am very grateful to Corentin, Marion, Quentin, Anjali, Anaïs, Victoria, Yann, Agata and Antoine for opening a path to a brighter collective future that I look forward to.

TABLE OF CONTENTS

List of tables	vi
List of figures	vii
List of symbols and abbreviations.....	ix
Chapter 1: Introduction.....	1
1.1. Research object	1
1.2. The role of data	2
1.2.1. The empirical foundation of phonological theories	2
1.2.2. My approach to data	3
1.3. Fostering inter-theoretical discussion.....	4
1.4. Approach and organization of the document	6
Chapter 2: Strong and weak syllables	7
2.1. Stress and reduced vowels.....	7
2.1.1. The vowel system of SSBE.....	7
2.1.2. Stress and accent	14
2.1.2.1. Terminological confusion?.....	14
2.1.2.2. Stress and feet.....	15
2.1.2.3. The indirect relation to vowel quality	16
2.1.2.4. The segmental diagnostics of stress	17
2.1.2.5. The interaction of stress and accent	22
2.1.3. Stress and segmental structure	26
2.2. Vowel reduction	29
2.2.1. First steps: The Arab Rule.....	29
2.2.2. The study of vowel reduction in pretonic syllables.....	31
2.2.2.1. Background	31
2.2.2.2. Methodology	32
2.2.2.3. Results	35
2.2.3. Vowel reduction in French loanwords	37
2.3. Pretonic secondary stress in simplex words	38
Chapter 3: The interaction between morphosyntax and phonology	45
3.1. Approach to morphology	45
3.1.1. Basic units	46
3.1.2. Morphological relatedness	51
3.2. How phonological theories incorporate morphosyntactic information.....	53
3.2.1. Prosodic Phonology.....	54
3.2.2. Stratal Phonology	56
3.2.3. Output-Output Correspondence	60
3.2.4. Strict CV.....	62

3.3.	Semantically opaque morphology and phonology	65
3.3.1.	Historically prefixed words	65
3.3.1.1.	Phonological evidence.....	67
3.3.1.2.	Morphological evidence	71
3.3.1.3.	Psycholinguistic evidence	72
3.3.1.4.	Semantic relics	74
3.3.2.	Complex proper names.....	76
3.3.3.	Learning to detect complexity in semantically opaque words	80
3.3.4.	The lexical and grammatical knowledge about non-compositional words	81
3.4.	Stress and syntactic categories	85
3.4.1.	Category-specific stress?.....	85
3.4.2.	Primary stress in verbs	87
3.4.2.1.	A comprehensive dictionary study	87
3.4.2.2.	A pseudo-word study on long verbs.....	96
3.4.3.	Primary stress in nouns	98
3.4.4.	Disyllabic pseudo-nouns and verbs.....	99
3.4.5.	Disyllabic verb-noun pairs	103
3.5.	Suffix classes.....	109
3.5.1.	SPE's boundaries.....	111
3.5.2.	The Guierrian approach.....	112
3.5.3.	Stratal models	114
3.5.4.	OO-Correspondence.....	118
3.5.5.	Two approaches to cohesion	121
3.5.5.1.	Raffelsiefen	121
3.5.5.2.	Newell	124
3.5.6.	Lexical conservatism.....	127
3.5.7.	Summary and perspectives	130
3.6.	Pretonic secondary stress in stress-shifted derivatives.....	135
3.7.	The segmentability hypothesis	137
3.7.1.	Hay's model	137
3.7.2.	Weak Preservation.....	139
3.7.2.1.	Exceptional stress preservation	139
3.7.2.2.	Stress preservation failure	142
3.7.2.3.	Pretonic vowel reduction in stress-shifted derivatives.....	145
3.7.3.	The gemination of [ɹ]	149
3.8.	Beyond locality and containment	157
3.8.1.	Distant bases.....	157
3.8.2.	Multiple parsings.....	161
3.8.3.	Faithfulness throughout the morphological family	161
Chapter 4: Orthography and phonology		167
4.1.	The previous literature and first explorations.....	167
4.2.	Orthography is not a side issue	169

4.3.	The limits of phonology	170
4.3.1.	‘Unnatural’ alternations.....	170
4.3.2.	Alternations requiring information provided by orthography	171
4.4.	Where does orthography stand?	173
4.5.	From secondary stressed vowels to a possible category-specific vocalic behaviour ..	177
4.5.1.	The value of secondary stressed vowels	177
4.5.2.	A foot-based perspective	179
4.5.3.	Vowel values and syntactic categories	181
Chapter 5: Sublexica in English.....		183
5.1.	Is there a single phonological system in English?.....	183
5.2.	Sketching a model	184
5.3.	Exploring the model	189
5.3.1.	The French sublexicon	189
5.3.2.	Final nuclei and sublexica	193
5.3.2.1.	The CMU data	193
5.3.2.2.	The experimental data	197
Chapter 6: Conclusion		200
References		202
Document de synthèse.....		223
Curriculum Vitae.....		250

LIST OF TABLES

Table 1. Five properties of the three series of vowels according to Lindsey (2019)	9
Table 2. Examples of differences in subsidiary prominence between the American and the British tradition	15
Table 3. Vowels and stress in different vocalic systems	16
Table 4. The distribution of /h/, /ɹ/ and of the allophones of /t/ in different varieties of English (after J. Harris 2013)	17
Table 5. The segmental diagnostics of stress proposed in the literature for English	22
Table 6. The factors claimed to affect vowel reduction	32
Table 7. Word counts in the different datasets of the study on vowel reduction	33
Table 8. Predictions of the position of pretonic secondary stresses according to the moraic trochee approach and the results observed. Brackets represent foot boundaries here.	42
Table 9. The two types of morphosyntactic conditioning acknowledged throughout the history of generative phonology (adapted from Bermúdez-Otero 2011)	54
Table 10. Predictions made by the presence or absence of the initial CV (T stands for an obstruent and R for a liquid)	64
Table 11. The phonological characteristics investigated in Dabouis (2023d)	77
Table 12. The phonological characteristics of opaque and transparent prefixed words	84
Table 13. Empirical studies of noun and verb stress in English.....	86
Table 14. Morphological distribution in the dataset used in Dabouis & J.-M. Fournier (2023)	89
Table 15. Weight coding adopted for deradical and simplex verbs in Dabouis & J.-M. Fournier (2023). Heavy syllables are shown in grey.	91
Table 16. Examples of stimuli used in the pseudo-verbs production experiment.....	96
Table 17. Test words and their different pronunciations used in the study initially conducted for Youssef Fadel's master's thesis	100
Table 18. Cases of overapplication of phonological processes induced by Class 2 affixation (H90 indicates that the process is described in J. Harris (1990) and B93 that it is described in Borowsky (1993))	116
Table 19. The properties of Class 1 and Class 2 affixes in early generative phonology according to Raffelsiefen (1999).....	117
Table 20. Probability of the /σ̀σ́(-)/ pattern in the models compared to the distributions found in the data	141
Table 21. Vowel features based on Jensen (2022)	146
Table 22. The pronunciation of <i>irregular</i> according to four pronunciation dictionaries.....	150
Table 23. Composition of the dataset in Dabouis <i>et al.</i> (submitted).....	152
Table 24. Summary of the properties of the different sublexica in English (adapted from Dabouis & P. Fournier (2022: 236))	186
Table 25. Criteria used for the coding of sublexica in the CMU data.....	194
Table 26. Statistical tests comparing the distribution of the data in the CMU and in the results from Experiment 2	198

LIST OF FIGURES

Figure 1. Vowel quality shifts which have taken place in standard British pronunciation (taken from Lindsey 2019: 18).....	8
Figure 2. Nature of the vowel of the second syllable depending on the weight of the first syllable in disyllabic words in the context of the “Arab Rule”	31
Figure 3. Vowels found for monographs in non-derived words depending on syllable structure	35
Figure 4. Vowels found for non-derived words depending on syllable structure and grapheme, among <a, e, o, u>	36
Figure 5. Proportions of words with non-final primary stress which may have a full vowel in their last syllable.....	37
Figure 6. Type of vowel found in the second syllable of non-derived words with two pretonic syllables.....	39
Figure 7. Position of secondary stresses in monomorphemic or deradical words with three pretonic syllables.....	41
Figure 8. Types of configurations with different associations of form and meaning (units that are typically assumed to be morphemic are shown in dark blue, those that are controversial are shown in green, and those that are usually not assumed to be part of morphology are shown in light orange)	49
Figure 9. Illustration of the terminology used to refer to morphologically related words	52
Figure 10. Stratal Phonology grammar	57
Figure 11. OO-correspondence (after Collie 2007: 232)	61
Figure 12. Position of primary stress in disyllabic nouns, adjectives, adverbs and verbs according to Guierre (1979)	67
Figure 13. Nature of vowels found in the initial pretonic syllable of prefixed and non-prefixed words that are stress-shifted derivatives (‘derived’) or monomorphemic words or words formed of a bound root and a suffix (‘non-derived’), in open or closed syllables (based on Dabouis & J.-M. Fournier to appear).....	69
Figure 14. Position of main stress in verbal compounds depending on the syntactic category of the first constituent and spelling (taken from Dabouis & J.-M. Fournier 2023).....	90
Figure 15. Position of primary stress in disyllabic deradical or simplex verbs depending on the weight of their final syllable (L = Light, H = Heavy), taken from Dabouis & J.-M. Fournier (2023)	92
Figure 16. Conditional inference tree for primary stress in deradical and simplex trisyllabic verbs (data from Dabouis & J.-M. Fournier 2023)	95
Figure 17. Conditional inference tree for primary stress in trisyllabic pseudo-verbs	97
Figure 18. Illustration of the questionnaire used in the study	101
Figure 19. Effect plot for the best regression model for the preferred position of stress in disyllabic pseudowords	102
Figure 20. Stress profile of prefixed verb-noun pairs depending on their relative frequency in British English, based on Wells (2008).....	105

Figure 21. Stress profile of prefixed verb-noun pairs depending on the weight of their final syllable in British English	106
Figure 22. Stress profile of prefixed verb-noun pairs depending on their date of first attestation in the <i>OED</i>	106
Figure 23. Position of primary stress in disyllabic verbs which have a homographic noun depending on different types of morphological structures	107
Figure 24. Conditional inference tree for position of primary stress in disyllabic verbs with a homographic noun. The predictors are morphological structure, weight of the final syllable and semantic dependency.....	108
Figure 25. Proportions of ‘stress shift’ in novel <i>-able</i> derivatives with a remote base in Breiss’s experiments depending on the weight of the syllable to which stress is shifted	130
Figure 26. Schematized dual-route model after Hay (2001). The solid line represents the decomposed route and the dashed line represents the direct route. Resting activation levels are represented by the thickness of the circles (<i>BNC</i> frequencies: <i>sane</i> (289), <i>insane</i> (360))....	138
Figure 27. Partial effects plot for the prefixed subset (n = 616; other significant predictors include derivative frequency).....	144
Figure 28. Vowels found for monographs <a, e, o, u> in non-prefixed derived and non-derived words, for both positions and depending on syllable structure	148
Figure 29. Vowels found for monographs <a, e, o, u> in derived and non-derived words containing an opaque prefix, for both positions and depending on syllable structure	148
Figure 30. Example of duration measurement for <i>irrelevant</i> (USM1)	153
Figure 31. Partial effect plot for the linear regression model on [ɪ] duration in the full dataset (taken from Dabouis <i>et al.</i> submitted).....	154
Figure 32. Effect plot for the linear regression model on [ɪ] duration in the derived words (taken from Dabouis <i>et al.</i> submitted).....	155
Figure 33. Example of geminated [ɪ] in British English (<i>irrelevant</i> - UKM23).....	156
Figure 34. The four possible routes of lexical access with a distant base.....	159
Figure 35. Partial effect plot for the subset without transparent prefixes (N = 1087)	160
Figure 36. Number of words in morphological family per derivative in Dabouis (2023b) ...	163
Figure 37. The BiPhon model extend so as to include orthographic representations (Hamann & Colombo 2017).....	174
Figure 38. The sublexica of English (taken from Dabouis & P. Fournier (2022: 248))	187
Figure 39. Conditional inference tree for the position of primary stress in French loanwords depending on syllable count, complexity and weight of the final syllable (in moras) in BrE191	
Figure 40. Conditional inference tree for the position of primary stress in French loanwords depending on syllable count, complexity and weight of the final syllable (in moras) in AmE	191
Figure 41. Comparison of CMU data in Moore-Cantwell (top) with our recoded data (bottom)	195
Figure 42. Conditional inference tree analysis of the CMU data.....	196
Figure 43. The distribution of sublexica in words ending in /ə/ and /i/ (with no final coda) in the CMU data	197
Figure 44. The distribution of the data in the CMU data and in Moore-Cantwell’s (2020) Experiment 2	198

LIST OF SYMBOLS AND ABBREVIATIONS

Acronyms

<i>SPE</i> :	<i>The Sound Pattern of English</i> (Chomsky & Halle 1968)
<i>BNC</i> :	<i>British National Corpus</i>
<i>COCAE</i> :	<i>Corpus of Contemporary American English</i>
<i>OED</i> :	<i>Oxford English Dictionary</i>
OT:	Optimality Theory
ASR:	Alternating Stress Rule (<i>SPE</i>)
SSBE:	Standard Southern British English
ERSaF:	English Root Stress across Frameworks (ANR-DFG project)
LLL:	Laboratoire Ligérien de Linguistique
LRL:	Laboratoire de Recherche sur le Langage
MFM:	Manchester Phonology Meeting (conference)
PAC:	Phonologie de l'Anglais Contemporain (project and conference)
RFP:	Réseau Français de Phonologie (network and conference)
ALOES:	Association des anglicistes pour les études de langue orale dans l'enseignement supérieur, secondaire et primaire

Vowel values

\check{V} :	Checked vowel
\bar{V} :	Free vowel
\bar{V}^f :	Foreign free vowel

Prosodic Phonology

Σ :	Foot
ω :	Phonological word
σ :	Syllable
μ :	Mora
X° :	Minimal projection of prosodic category X
X' :	Higher projection of prosodic category X

Levels of representation

[...]:	Phonetic representation
/.../:	Phonological representation
... :	Underlying representation (also used for elements in Element Theory)

Other symbols

§:	Sublexicon
◆:	Morphologically related words
/1/ or ' or ´:	Primary stress
/2/ or , or `:	Secondary stress
{a, b, ...}:	List of possible elements

CHAPTER 1

Introduction

I was introduced to phonology during my studies at the University of Tours, in my third year of the LLCE bachelor's degree. The third-year phonology course taught by Jean-Michel Fournier is the main element that decided me to enrol in a linguistics master's degree, in which I would make my first steps as a researcher. During that degree, I discovered the Guierrian School, its history, philosophy and results, along with other phonological theories – notably Optimality Theory and Prosodic Phonology, in a course on language acquisition taught by Christophe dos Santos. My education as a linguist owes a lot to the teachers of other disciplines who taught in that degree : Sylvester Osu, Philippe Prévost, Sylvain Gatelais, Fabienne Toupin and Laurie Tuller. They gave me a broader view of what studying language can be and I am lucky to have been able to follow their courses. I was also soon integrated into the dynamic morphophonology research group of the Laboratoire Ligérien de Linguistique (henceforth LLL) and was able to take part in their work on stress in disyllabic verb-noun pairs. I also got to attend the PhD defences of several PhD students who were part of that group, including those of Pierre Fournier and Marjolaine Martin, and to be invited to join the group in international conferences. From my very first year as a researcher, I learned two things that have been key features of my work up to this day: the importance of the **empirical foundations** of any theoretical model and the importance of being **open to different theoretical frameworks**. I will expose them in more detail in the rest of this introductory chapter before concluding with a presentation of the general organisation of this document. However, before getting into this, let me say a word about what I see as being the main object of my research.

1.1. Research object

My work has focused on **the phonology of words** (or lexical phonology) in (mostly British) English: the presence and distribution of lexical stresses and the distribution of segments. I have been interested in the mechanisms that determine those distributions, from the more clearly phonological ones to those that involve interactions with other linguistic 'levels' or 'modules' such as morphosyntax or orthography. This issue of word level phonology is the main general issue that I am trying to understand, in itself but also as a window into other phenomena (e.g. the organization of the mental lexicon). Thus, I have not been working on what happens to words when they are inserted into sentences, and how the context into which words are found may affect their phonology, although I do believe that the relationship between the prototypical realization of words and the context into which they usually occur would be a fruitful area of research. For example, Kelly & Bock (1988) argue that the rhythmic contexts in which

disyllabic nouns and verbs appear might contribute to the differences in stress placement between those two syntactic categories: nouns appear more often in rhythmic contexts that favour trochaic stress (i.e. a preceding unstressed syllable and a following stressed syllable) while verbs appear more often in contexts favouring iambic stress (i.e. a preceding stressed syllable and a following unstressed syllable).¹ Another example might be antonymic pairs such as *demote* ~ *promote*, *exterior* ~ *interior* or *inhale* ~ *exhale*, for which I found in my dissertation that the member of the pair with the lowest frequency tends to have secondary stress on its prefix while the more frequent one does not. It is possible that the presence of this secondary stress is a lexicalized contrastive stress, which is used to distinguish the member of the pair which is less frequent from the more frequent member, which is more ‘unmarked’ and ‘expected’. Those two short examples illustrate that contextual effects certainly interact with word level phonology and the details of how that happens should be investigated further.

As will be seen in the following sections, I have been working on a variety of different issues, and one may not see what ties everything together at first sight. On top of the two key features mentioned previously, another important characteristic of my approach to research has been the attention given to the **overall system**. I have always looked at the different processes that I have worked on as pieces of a very complicated puzzle. Chomsky & Halle’s *Sound Pattern of English* (henceforth *SPE*) was an attempt to give a general account of the phonology of English, and so was Guierre’s (1979) study of stress patterns and spelling-to-sound correspondences. Inspired by those endeavours, I tend to approach the study of specific processes with what I know about the rest of the system in mind. In this document, I have showed the multiple connections that exist between the different topics that are covered by regularly cross-referencing sections.

1.2. The role of data

1.2.1. *The empirical foundation of phonological theories*

The way that I view the relationship between data and theory has been strongly influenced by the way that the approach introduced by Guierre has defined itself in opposition to the way that generative phonology has often been practiced up until the 2000s. In *SPE* and many publications that it inspired, it is often unclear what the source of the data is, and we are usually not told how many words are concerned by a given rule and how many exceptions there are. We get data based on introspection, and sometimes quite exotic words are cited as illustrations of the phonology of English (e.g. *Monongahela*, *Apalachicola*, *Ticonderoga*).² By contrast, one of the key characteristics of the Guierrian School has been the use of large computerized and annotated datasets based on the transcriptions of pronunciation dictionaries.³ Guierre (1979)

¹ The methodology that they use has some limitations, which are discussed in Dabouis & J.-M. Fournier (2023).

² Those sometimes get repeated in dozens of papers and, strikingly, one of those, *hamamelidanthemum*, is exclusively found in phonology papers and does not appear to have any real usage by non-linguist users of English. This general problem of excessive reliance on introspection and stock examples has been noted by others closer to the post-*SPE* tradition, e.g. Collie (2007: 3), McMahon (2001) and Wenszky (2004: 12).

³ This approach was founded by Lionel Guierre (1979), who was the first professor of English phonetics and phonology in France (see the obituary written by Deschamps & O’Neil (2007)). It can be defined through a combination of five main characteristics: the use of large datasets based on pronunciation dictionaries, the

argued for what he called a “probabilistic phonology”, in which “probabilistic” should not be taken to mean the same thing as in probabilistic models of phonological computation (some of which will be discussed further on). By that, Guierre meant that rules should be tested against large datasets, so that we should be able to say how many words are concerned by a given rule, how many exceptions there are and what the efficiency of the rule is. Those pieces of information are of theoretical interest, as he argues that more numerous and more regular classes exert a stronger “statistical pressure” on their members, and that this makes those members more predictable.

With the expansion of corpus linguistics and the facilitated access to large datasets, the use of introspection and stock examples taken over from one publication to the next has somehow diminished, although it has not disappeared entirely (see Jensen (2022) for a recent example). However, as will be seen in several sections of this document, there are many areas in which the empirical foundations of certain claims are very narrow and fragmentary.⁴ This means that some of those claims might rest on quite fragile ground, and that much work remains to be done for us to have greater confidence in them.

1.2.2. *My approach to data*

Having been trained in the Guierrian School, I have started by conducting large-scale studies based on data from pronunciation dictionaries, and I still use that source of data. Dictionaries have well-known drawbacks (e.g. they mix phonemic and phonetic information (Dahak 2006), refer to an “artificial idiolect” (Collie 2007), provide syllable boundaries that are based on theoretical grounds and thus differ from one dictionary to the next (Ballier & P. Martin 2010), and they do not list all attested variants (Ballier 2016)) – and less often mentioned advantages (e.g. access to large numbers of words, idiolect uniformity, no use of automatic rules in their conception). I see them as simply one source of data among many others, which all have their drawbacks and advantages. Ideally, I believe that we should triangulate the processes that we are interested in by using different types of data – whenever relevant and possible. Following some of what Ballier (2016) calls the “second generation” of the Guierrian School (M. Martin 2011; Videau 2013), I have been going beyond dictionaries, and so I will here report several studies that use other types of data (e.g. speech data, experimental data). The ‘triangulation’ that I have just mentioned is one the research strategies developed in the context of the ERSaF (“English Root Stress across Frameworks”) project, which is a project which I am in charge of along with Sabine Arndt-Lappe and in which we co-supervise two PhD students, Aaron Seiler and Marie Gabillet.⁵ Some of the first results that have come out of this project will be discussed in this document.

inclusion of orthographic information in phonology (see Ch. 4), an approach to morphology which includes semantically opaque formatives as potentially relevant to the phonology (see §3.3), the fact that the role of syllable structure in stress assignment is viewed as minor (especially in comparison with post-*SPE* approaches; see §2.1.3) and the assumption that English phonology is not monolithic, so that different sublexica have slightly different phonological grammars (see Ch. 5). For a general presentation of the approach, see Dabouis *et al.* (2023).

⁴ The most striking cases are the segmental correlates of stress (§2.1.2.4) and the segmental processes that distinguish different suffix classes (§3.5).

⁵ For a description of the project and its first communications, see: https://lrl.uca.fr/projet_du_lab/projet-ersaf/ or <https://www.uni-trier.de/universitaet/fachbereiche-faecher/fachbereich-ii/faecher/anglistik/departamental-branches-fachteile/linguistics/-/sprachwissenschaft/sabine-arndt-lappe/arndt-lappe-projekt-ersaf>

Ballier (2016) describes the Guierrian School as a form of corpus phonology. Indeed, some colleagues whose work can be characterized as Guierrian seem to embrace the view that the data is what matters, and sometimes strongly reject more formal approaches. My personal position sits in between the two. On the one hand, I see the value in using large corpora, especially considering the importance that introspection and intuition have had in studies on English phonology. The question of ‘how we know what we know’ about English phonology is a central concern of mine. Thus, the use of large datasets is a recurrent characteristic of my research.⁶ However, I am quite aware of the limitations of corpora: they cannot tell us what is impossible, they are not an optimal tool to study rare processes and they are not always large enough to answer a given question (Ferragne 2021: §2.3; Scheer 2013). Moreover, using large corpora is a way to do research (which is decreasingly remarkable), and it cannot entirely define a research program as it is not an object of study in itself. As Popper (2002 [1959]: 7) puts it, “a hypothesis can only be empirically *tested* – and only *after* it has been advanced” (emphases in the original). On the other hand, I take theories seriously (see the next section), and they provide a framework in which we can think about phonology, questions to answer and predictions to test. Thus, the issues that are at the core of many of the studies that I will be discussing in this document have come out of exchanges with theoretical frameworks which are not the one that I was trained in.

Let me conclude this section with a word on the tools that I have been using. Following Guierre, much of the Guierrian literature only makes use of basic descriptive statistics (raw figures, percentages) on manually annotated datasets based on pronunciation dictionaries. This is how I started, but quite quickly I added visual representations of the distributions of the data, which I believe to be considerably more informative than figures in a table, and so I completely subscribe to Ferragne’s call for a “culture of the graph” (Ferragne 2021: 21-22). While I was preparing my dissertation, I followed an introductory course to statistics using the software R. This is a tool that I have never abandoned and have been increasingly using, even though I am convinced that I have but scratched the surface of what it can do. Therefore, I have been using inferential statistics in many of the recent studies that I have been conducting. While one may be annoyed when a reviewer asks for a chi-square test when the raw figures obviously show that the difference between two subsets of the data is strongly significant, the use of multivariate regression models has been particularly necessary to approach processes which are highly variable and involve many different predictors. More recently, I have been using Praat (Boersma & Weenik 2024) to study the duration of [ɪ] in prefixed words in *ir-* (see §3.7.3).

1.3. Fostering inter-theoretical discussion

When reflecting upon the relationship that phonological theories entertain with one another, a term that I have encountered quite recently comes to mind: schismogenesis.⁷ The term describes how social groups can define themselves in opposition to another group, sometimes in many

⁶ I am also increasingly convinced of the virtues of open data, and so I have been making my datasets available on OSF: <https://osf.io/a82nh/>

⁷ I found out about this concept when reading the fascinating *The Dawn of Everything: A New History of Humanity* of Graeber & Wengrow (2021) on the history of early human societies.

different ways. I mentioned above that certain Guierrian researchers strongly rejected what they see as excessive formalism in other theories. Among more widespread generative phonological theoretical frameworks, we also find strong divisions between researchers who adopt Optimality Theory (henceforth OT; Prince & Smolensky 1993) as the computational device, who use Prosodic Phonology and assume that subsegmental units are binary features, and researchers of the Government Phonology tradition, which tend to use Rule-Based Phonology for computation, reject the arboreal representations of Prosodic Phonology and prefer to use lateral relations between consonants and vowels instead, and assume that subsegmental units are elements. While there are some cases of blends of those two traditions (e.g. Faust & Torres-Tamarit 2017), they tend to evolve relatively independently from one another. The comparison of theories is complicated by the fact that they tend to work on different issues and disagree on the basic units of phonology, and even on the very nature of what phonology is. Moreover, the available evidence itself sometimes makes the choice between different theories very difficult (this is called analytic underdetermination, see Bermúdez-Otero 2012: §2).

When I entered the world of research in phonology, I was initially myself quite critical of those theories that seemed overly formal and sometimes relied on such small datasets to develop their analyses. However, from the start it seemed necessary to learn about them, their preoccupations, their vocabularies and their predictions to be able to exchange with them. Because of the difficulty of the patterns found in human language and, regarding what I know best, English phonology, I have never wanted to commit to a particular theory, as I do not believe that any existing theory can pretend that it has absolute superiority over others. However, I see theories as very useful tools for several purposes. First, they can help us make sense of phenomena that are quite complex. Certainly, nowadays the increasingly powerful so-called ‘artificial intelligences’ can predict certain processes with an accuracy that outshines many linguistic models, but they cannot answer our desire to understand why things are the way they are and how they function. Second, theories make predictions which can lead us to look for things in the data that we would never have thought to look for (e.g. in physics, Neptune or gravitational waves would not have been discovered without theories which predicted their existence).

Therefore, I have been using tools from different theories, explored their predictions and sought to put them to the test. That will to compare and test the predictions made by different theoretical frameworks is at the core of the ERSaF project. Moreover, I have embraced the idea that we do not improve our understanding of the world by throwing away whole theories because they fail to account for some things that we deem important, but rather that we incrementally adapt them to new data if it contradicts their predictions so as to broaden their empirical coverage. On several issues, I have also found useful to go beyond the phonological literature and look for what the evidence is regarding those issues in the psycholinguistic literature or the literature on acoustic phonetics.

Finally, composing this document has been a new step in my understanding of existing models as it was the occasion to take more time to read about them. Thus, I have learned more about Government Phonology, various theories of morphology or the BiPhon model. I have tried to synthesize some of those readings here in the hope that they will be useful to those curious to learn about them.

1.4. Approach and organization of the document

In this last part of the introductory section, I would like to present the ‘spirit’ in which I prepared this document and how I chose to organize the different topics on which I have worked.

My approach has been to give a condensed overview of my work through the lens of the two central components of my research (extensive data and inter-theoretical discussion) and to critically present some of the studies that I took a part in (especially the older ones), and to propose new extensions of some of those studies. I have also integrated a few elements which have emerged in the year that I have devoted to this synthesis and have not been published yet. These will be of a more exploratory nature than the rest of the document. Moreover, to make the document more accessible to readers who are not entirely familiar with all the theoretical concepts that are relevant to the studies that I will be discussing, I have included sections which introduce those concepts and which are not directly based on my own work.

Regarding the overall organization of the document, it has been a challenge to decide how to arrange the topics that I have worked on in a coherent and not overly redundant way. Many of the studies that I have conducted are relevant to different issues in English phonology and this led me to split certain studies into different parts, in the sections that were most relevant to those parts. I see this document as a broad overview of the issues that I have been interested in regarding the phonology of English, without aiming at being an exhaustive overview of all the issues of the word level phonology of English. However, I believe that anyone who is interested in the phonology of English will find here many key issues and – I hope – useful summaries of those issues: the nature of the rhythmic organization of English and the relationship between the segmental and the suprasegmental (Ch. 2), the nature of morphology and how it interacts with the phonology (Ch. 3), the role of orthography in phonology (Ch. 4), and the broad divisions in the lexicon (Ch. 5).

The following chapters will refer to the results of many statistical analyses and empirical investigations. To make the document more reader-friendly, the details of those tests (especially complex regression analyses) and the details of the methodology of each study will be left out and the explanation of the results and graphs will be favoured. The reader who is interested in the details of those statistical tests and methodologies is invited to consult the published studies for the full details.

CHAPTER 2

STRONG AND WEAK SYLLABLES

A central part of my research has focused on the distribution of strong syllables, and so it seems important to start by laying out my current view of what stress and accent are (§2.1). In this first section, I will start by discussing the nature of the vowel set in Standard Southern British English (henceforth SSBE), which will be a new contribution that has not been presented or published elsewhere. I will also discuss the nature of stress and accent (§2.1.2) and the relationship between stress and segmental structure (§2.1.3). I will then review some of my work on so-called ‘vowel reduction’ (§2.2) and pretonic secondary stress (§2.3) when they do not interact with morphology.

2.1. Stress and reduced vowels

2.1.1. *The vowel system of SSBE*

Like many phonologists and phoneticians working on English, I have been using the transcription conventions developed in the British tradition, which are represented in the two most prominent pronunciation dictionaries, Jones (2006) and Wells (2008). If the symbols that are used to describe the vowels of SSBE are to be in line with the International Phonetic Alphabet, then it seems like the variety that we are transcribing is rather old-fashioned. Indeed, it is now clear that British speakers do not pronounce TRAP words⁸ with [æ] but rather with a more open [a], that the vowel in DRESS has lowered from [e] to [ɛ] or that the vowel in SQUARE has now become a long monophthong [ɛ:]. This growing gap between the symbols and present-day pronunciations have led to a mini-reform in the French agrégation so as to accept new symbols reflecting some of those recent evolutions. More recent dictionaries such as Upton & Kretzschmar (2017) or the updated transcriptions on the online *Oxford English Dictionary* (henceforth *OED*) also reflect some of those changes. However, there are more radical approaches that argue for a complete reanalysis of the vowel system of SSBE. One prominent defendant of such a radical approach is Lindsey (2019). Having myself felt uneasy by this growing gap between the old symbols and the current realizations of vowels, this habilitation was the occasion for me to explore the alternative proposal made by Lindsey (and others).

Lindsey claims that diphthongs are better analysed as **vowel-glide sequences** ending either in /j/ or /w/, and that there are seven diphthongs in SSBE.⁹ He strongly criticizes the symbols that are commonly used for FLEECE and GOOSE, /i:/ and /u:/, and argues that those two

⁸ Small capitals are used here to refer to lexical sets after Wells (1982).

⁹ Lindsey reports many other changes that are not discussed in detail here, such as the backing of FOOT, GOOSE, CURE and GOAT before [ɪ] or the mergers of THOUGHT, GOAT and LOT before [ɪ] and KIT and FOOT with schwa.

vowels are diphthongs which pattern with other diphthongs such as FACE or MOUTH, and not with long vowels such as NURSE or SQUARE. The vowel inventories for RP (according to Wells (1982, 2008)) and SSBE (according to Lindsey (2019)) are shown (1).

(1) The vowel inventories of RP and SSBE¹⁰

	Short vowels		Diphthongs		Long vowels			
	RP	SSBE	RP	SSBE	RP	SSBE		
KIT	ɪ	ɪ	FLEECE	i:	ɪj	NEAR	ɪə	ɪ:
DRESS	e	ɛ	FACE	eɪ	ɛj	SQUARE	eə	ɛ:
TRAP	æ	ɑ	PRICE	aɪ	ɑj	START	ɑ:	ɑ:
STRUT	ʌ	ə	MOUTH	aʊ	aw	NURSE	ɜ:	ə:
LOT	ɒ	ɔ	CHOICE	ɔɪ	oj	FORCE	ɔ:	o:
FOOT	ʊ	ə	GOAT	əʊ	əw	CURE	ʊə	ə:
			GOOSE	u:	ʊw	LETTER	ə	ə

Beyond the reanalysis of diphthongs as vowel-glide sequences, the symbols used in (1) reflect the “anti-clockwise” shift that has happened between RP and SSBE, as shown in Figure 1.

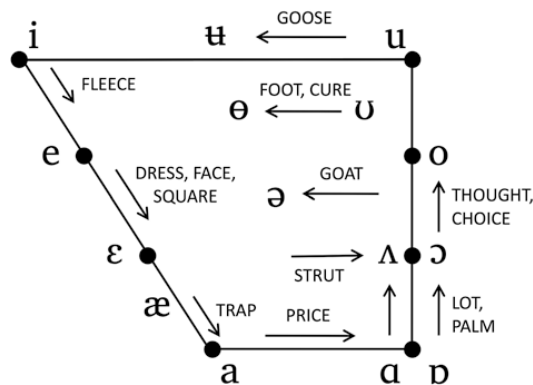


Figure 1. Vowel quality shifts which have taken place in standard British pronunciation (taken from Lindsey 2019: 18)

A few notes on this figure should be made. First, one can note that STRUT is not here analysed as /ə/, and Lindsey (2019) notes that this vowel has moved from a central position to a more back position. THOUGHT has also known raising, although the figure probably simplifies the situation, and although the symbol to describe this vowel was /ɔ:/, it was probably not as low as the sixth cardinal in RP. Likewise, DRESS was not as high as the second cardinal. As noted by Lindsey (2019: 45) and pointed out to me by Péter Szigetvári (p.c.), MOUTH is moving in the opposite direction: while its starting point used to be close to [a], it is now more front and closer to [a].

¹⁰ Note that /ə/ appears in two sets of vowels, as STRUT and LETTER. Szigetvári (2018) assumes that they are the same phoneme, and that they only differ in stress (STRUT is stressed while LETTER is its stressless realization). CURE is shown in grey as it often merges with FORCE (Cruttenden 2014: 156; Lindsey 2019: 48).

Lindsey (2019: 24, 50) argues that the three series of vowels in (1) differ from one another based on five properties, as shown in Table 1, to which we add two environments mentioned in Szigetvári (2016). Crucially for the analysis of FLEECE and GOOSE, those two vowels pattern with the other diphthongs.

		Short vowels	Diphthongs	Long vowels
May be followed by	V	×	✓	×
	#	×	✓	✓
	C	✓	✓	✓
	linking-r	×	×	✓
Particularly sensitive to pre-fortis clipping		×	✓	×
May trigger pre-liquid breaking		×	✓	×
Undergo smoothing before [ə]		×	✓	×

Table 1. Seven properties of the three series of vowels according to Lindsey (2019)

The first four properties in Table 1 refer to **distributional properties** which are quite well established. The vowels triggering linking-r are quite recognized as being those which may be spelled with an <r>, but occasionally may be spelled otherwise (the process is then usually called ‘r-intrusion’). As for vowels that occur in prevocalic position, one needs to assume that intrusive-r is always realized, which we know is not the case (see Durand *et al.* 2015; Mompeán-Gonzalez & Mompeán-Guillamón 2009), and so vowels such as SQUARE or NORTH may be followed by another vowel. However, the generalization holds if we restrict its scope to certain items: monomorphemic words, words derived from bound roots and words containing stem-level/cohering affixes. An alternative might be to state that those are the vowels which can occur prevocalically without triggering any sort of ‘repair’.

The other three properties put forward by Lindsey have weaker empirical foundations. Let us first consider **pre-fortis clipping**. While it is quite well established that English vowel duration is affected by the voicing of the following consonant, those followed by a voiceless (‘fortis’; e.g. *bat*) consonant being shorter than those followed by a voiced (‘lenis’; e.g. *bad*) consonant (M. Chen 1970; Gonet & Stadnicka 2005; van Santen 1992), it is very difficult to find peer-reviewed evidence that supports the idea that diphthongs are ‘clipped’ more than monophthongs. The studies that I have been able to find have not tested that factor. Tanner *et al.* (2019, 2020) study a very large corpus of oral data from various corpora representing different varieties of English in the British Isles and North America, but they do not factor in the nature of the vowel in any way. Morley & Smith (2023) do test “vowel class” but their classification distinguishes short vowels from long vowels, and the latter category includes both diphthongs and long monophthongs. It is thus not possible to see if there are any specificities for diphthongs. Coretta (2020) presents the results of a production study, in which monosyllabic and disyllabic pseudowords (e.g. *teep*, *terkus*) were read in frame sentences. The pseudowords contain only three different vowels, /ɪj/, /ə:/ or /ɑ:/ (all analysed as long vowels, mainly differing in vowel height). His results show that /ɪj/ is significantly more affected by the voicing of the following consonant than the other two vowels, which are not different from one another, which is thus consistent with Lindsey’s claim. Lindsey (p.c.) recognizes that there is indeed a lack of

evidence on the matter, although trained phoneticians “can hear the clipping differences”. Nonetheless, he conducted a small study on audio samples mostly from pronunciation dictionaries which he shared with me. Although the sample is small and the type of speech is far from natural, the results show no statistically significant differences in the amount of clipping found among short and long monophthongs, but significant differences between diphthongs and both types of monophthongs. He finds that pre-fortis diphthongs are 40-50% of the duration of pre-lenis ones, while pre-fortis monophthongs are 70-80% of the duration pre-lenis ones. Thus, we can conclude that there is insufficient evidence to strongly support Lindsey’s claim on the greater sensitivity of diphthongs to pre-fortis clipping as compared to monophthongs, but that the little available evidence is consistent with his claim.

Pre-liquid breaking is the process by which a schwa-like vowel can be heard between a vowel and a following [ɫ] or [ɹ] (e.g. *file* [ˈfajɫ]; *real* [ˈriɹɫ]). I was only able to find two phonetic studies on the matter (Gick & Wilson 2001; Riera & Romero 2007), which report that the process is highly variable, far from a discrete process of epenthesis but that the schwa-like sound found in those cases is rather a transition sound, and that it seems to affect all vowels alike. The rest of the available literature on the issue is made up of phonological analyses relying on data built from introspection or intuition. McCarthy (1991) reports that in Eastern Massachusetts, schwa epenthesis is “required” before /l/ (e.g. *fail* [ˈfeɪəl]) if the word is followed by a consonant or a pause and that this contrasts with words with an “underlying” schwa (e.g. *betrayal* [biɪˈtreɪəl]), which preserve it in all positions. Nothing is said as to where those observations come from. Lavoie & Cohn (1999) discuss what they call “sesquisyllables”, i.e. words whose syllable count oscillates between one and two, precisely because of possible pre-liquid breaking. Their data is their own intuitions and those of six North American English participants, and they find that low and short vowels are uncontroversially monosyllabic while diphthongs and tense vowels may be disyllabic (although there is some variation depending on the precise vowel and the nature of the liquid). Lindsey himself relies on Wells’s (2008) note on the issue, which indicates that breaking “may happen when l follows i:, eɪ, aɪ, ɔɪ [...] (Some speakers of GenAm have pre-l breaking after u:, ʊ, ʌ [...]”. Although not all diphthongs are concerned for SSBE according to Wells, the important point for Lindsey’s analysis is that FLEECE (and GOOSE for General American) patterns with other closing diphthongs.¹¹

Smoothing is a process by which the second part of a diphthong is lost when it precedes schwa (and sometimes /ɪ/) and, in SSBE generates long monophthongs (Lindsey 2019: 46), while in RP it generated centring diphthongs. I have not been able to find any phonetic study of the topic, but the process is mentioned by Wells (1982: 238 ff.), who notes that the process applies to FLEECE and GOOSE, which he says “can be interpreted as evidence in favour of analysing FLEECE and GOOSE as underlyingly diphthongal”. Cruttenden (2014: 150 ff.) also discusses the process in RP and says that “a similar weakening [...] sometimes” affects FLEECE and GOOSE. Lindsey provides audio samples illustrating how present-day smoothing functions on his blog¹², and summarizes it as in (3).

¹¹ Szigetvári (2016) suggests that homorganicity between /w/ and /l/ is what could explain the greater stability of those clusters (i.e. the more common absence of breaking) as compared to /j/ and /l/.

¹² URL: <https://www.englishspeechservices.com/blog/smoothing-then-and-now/> [Consulted 04/10/2023].

(2)	ɪj + ə	→	ɪ:	ɥw + ə	→	ɥ:/ə:
	ɛj + ə	→	ɛ:	aw + ə	→	a:
	ɑj + ə	→	ɑ:	əw + ə	→	ə:
	oj + ə	→	o:			

In one of his YouTube videos¹³, Lindsey puts forward another argument that does not appear in Table 1. He challenges the claim that there is a process of “**glide-insertion**” after diphthongs, contrary to what is commonly found in textbooks. This process would consist in inserting a glide after a diphthong when it is followed by a vowel (e.g. *try* [j] *out*, *how* [w] *old*), and would be triggered to avoid a hiatus. In that regard, the process would be comparable to linking-r, which is another process assumed to be some kind of hiatus breaker. Lindsey argues that, if this analysis is correct, we would predict *two o'clock*, *three o'clock* and *four o'clock* to have the same duration, as they would all involve an initial word with a long vowel (/u:/, /i:/ and /ɔ:/) followed by an epenthetic consonant inserted to break the hiatus. In his video, he reports results from a mini-study in which he extracted audio extracts for those expressions “in normal running speech videos in YouTube videos” (15 speakers for British English) and measured the duration of the sequence starting from the vowel of the first word to the /ə/ starting *o'clock*. He reports a significant difference between *two o'clock*, *three o'clock* on the one hand and *four o'clock* on the other, the latter being longer than the former two. This would be evidence that there is only an inserted sound in *four o'clock* while the glides [w] and [j] in *two o'clock* and *three o'clock* are not inserted but merely resyllabified. Lindsey compares this with comparable data for North American speakers, for whom there is no significant durational difference between those three expressions, arguably because the vowel in *four* is not long, and the /ɹ/ is always present because of the rhotic character of that variety of English. Thus, Lindsey’s analysis for those two varieties can be summed up as in (3).

(3)	a. British English	b. American English
	<i>two o'clock</i> 'tʉw + ə'klɔk → 'tʉ wə'klɔk	'tʉw + ə'klɔk → 'tʉ wə'klɔk
	<i>three o'clock</i> 'θɹɪj + ə'klɔk → 'θɹɪ jə'klɔk	'θɹɪj + ə'klɔk → 'θɹɪ jə'klɔk
	<i>four o'clock</i> 'fo: + ə'klɔk → 'fo: ɹə'klɔk	'foɹ + ə'klɔk → 'fo ɹə'klɔk

However, this mini-study has not been published and we do not have the full details of how it was conducted. In order to be sure that the difference is indeed significant, one would have to replicate the study on a larger scale, collect considerably more data and normalize speech rate.

We have just seen that, although there is not much available evidence for those different processes, what is available broadly converge with the claims in Table 1, and it places FLEECE and GOOSE in the same category as other closing diphthongs. Analysing diphthongs as vowel-glide sequences is another matter and it is worth considering **what it can account for**.

The first thing that such an analysis can account for is the **distribution of diphthongs** shown in Table 1: they are the only ‘vowels’ which may occur prevocally because they do

¹³ Lindsey, Geoff, “LENGTH and LINKING in British, American and Australian accents!”, uploaded by Dr Geoff Lindsey, 15 October 2023, <https://www.youtube.com/watch?v=tPi2jtU7Tl4>

not end in a vowel. Therefore, a sequence such as /əjə/ does not need ‘repairing’ because it does not contain a hiatus. What is sometimes analysed as “glide-insertion” can simply be analysed as resyllabification¹⁴, with the glide moving from the coda to the onset of the following syllable. In the graphophonological model developed by J.-M. Fournier (2010b), there is a rule predicting that a stressed vowel should be free when it is followed by zero consonants (the C⁰ rule). This rule is the result of the unification of two rules identified by Guierre (1979), which noted the systematic presence of free vowels in two contexts: word-finally (e.g. *flý*, *gó*, *ré*) and prevocally (e.g. *líon*, *póet*, *muséum*). If we assume that diphthongs are vowel-glide sequences, then their presence in prevocalic position could be attributed to a strong constraint against hiatuses¹⁵, but the same explanation cannot be invoked for word-final vowels. One possible explanation for the presence of diphthongs word-finally is minimality requirements. In analyses of English phonology using Prosodic Phonology, it is assumed that a phonological word contains at least one foot, and that feet are minimally bimoraic (see §2.1.2.2). Thus, short vowels being monomoraic, they are banned from word-final position in stressed syllables, and so only diphthongs and long vowels may be attested in that position.¹⁶ Thus, it is possible that there are different phonological motivations for those two generalizations, so that J.-M. Fournier’s unification of those two rules is really to be seen as a graphophonological generalization.

J.-M. Fournier (p.c.) assumes that **smoothing** is attributable to hiatus avoidance: the second element of a diphthong is deleted and the two vowels merge into a single long vowel to repair a hiatus. If one analyses diphthongs as vowel-glide sequences, that analysis is no longer possible because there is no hiatus when a diphthong is followed by a vowel. Thus, another explanation should be provided for that process. As pointed out above, Cruttenden describes smoothing as a form of lenition (he uses the terms “weakening” or “reduction”), which we might conceive as a reduced constriction between two vowels which progressively leads to the loss of intervocalic glides. This is a rather common process, as noted by Picard (2003), who cites examples from Bourciez & Bourciez (1967: 1970), such as the loss of Latin /w/ (written <v>) between vowels in French (e.g. Lat. *pāvōnem* → Fr. *paon* ‘peacock’). In more conservative RP, smoothing usually resulted in centring diphthongs (e.g. *tire* /'tʰajə/ → [tʰaə]; Lindsey 2019: 46), but it now results in long monophthongs in SSBE. A possible analysis of this might be to assume that the process has been phonologized and that the loss of the medial

¹⁴ It is possibly that resyllabification from the coda to the onset which explains the clearer perception of glides. Following Raffelsiefen (2023), I will assume that the glides found in diphthongs syllabify in the coda as a default (although she does not assume that they are glides, see fn. 15). She does not assume that resyllabification may occur, though I assume that if there is such resyllabification, it may be post-lexically (and maybe, in a stratal model, at the word-level). More investigation of that assumption will be needed to establish in which conditions they might be syllabified otherwise.

¹⁵ Raffelsiefen (2023) assumes that this constraint is *NN (“no adjacent nuclei”) and she argues that, in General American, /i/ and /u/ associate to the nucleus and the coda. This would explain why they can be found prevocally. This alternative analysis is very close to the one presented here, and more detailed investigation would be required to establish in what they differ.

¹⁶ Péter Szigetvári (p.c.) suggests an alternative analysis that the absence of short vowels word-finally might be attributable to a phonotactic constraint *V#, which would leave only word-final unstressed schwa to account for. While this could account for the distribution of vowels word-finally, it could be seen as a loss of generalization as it cannot capture what the assumption that feet are minimally bimoraic can, such as the ‘Arab Rule’ (§2.2.1), Trochaic Shortening (§4.5.2) and other generalizations throughout the history of English exposed in Bermúdez-Otero (2018a).

element does not result in the two remaining ones anymore, but that the melody of the first element spreads to the second vocalic element. This would be consistent with analyses which assume that schwa has no melodic content (see e.g. J. Harris 2005). Note that the dispreference for hiatuses may be one of the forces pushing for a fusion of the two vowels once the glide has been lost.

Szigetvári (2016) analyses **pre-liquid breaking** and argues that if one analyses diphthongs as vowels, then “it is not easy to see what would force epenthesis between a vowel and a consonant”, while this is more straightforward if one assumes that they are vowel-glide sequences: an epenthetic [ə] is inserted to break a sequence of consonants that are too similar in sonority.¹⁷ He argues that that strategy has also been used to adapt loanwords which do not have a falling sonority contour word-finally (e.g. Fr. *mètre* /mɛtʁ/ → Eng. /'mɪjtəɪ/; Old Fr. *temple* → Eng. /'tɛmpəl/)¹⁸ or to break final clusters that are too similar to one another to form a cluster (e.g. *matches* /'matʃəz/).

He also puts forward another argument, which is the **distribution of flapping** in New Zealand English. In the New Zealand Acrolect, flapping appears to be only possible after short vowels (e.g. *matter*), while it is possible after a long monophthong (e.g. *martyr*) or a diphthong (e.g. *mitre*) in the Basilect. Szigetvári argues that, in the Acrolect, flapping may only occur intervocalically. This implies that a word like *martyr* somehow is still analysed as having an /ɪ/, even if the variety is non-rhotic. An alternative explanation may be proposed using foot structure. Under the assumptions that a short vowel necessarily forms a minimal foot with the following syllable (see §2.1.2.2) and that flapping targets /t/ and /d/ that are not foot-initial, we could assume that the Acrolect restricts flapping to minimal feet, while other varieties may have broader domains of application for that process.

Let us also consider a possible counterargument put forward by Polgárdi (2015). In verbs, it is often assumed that primary stress is final if the final syllable is heavy and penultimate otherwise, but that the final consonant does not contribute weight (see §3.4.1). Therefore, if diphthongs are vowel-glide sequences, then the glide would not contribute weight word-finally and so we would expect verbs that end in diphthongs to have penultimate stress. She argues that this is not the case (e.g. *deny* /də'naɪ/, *agree* /ə'griː/). As will be seen in §3.3.1.1, a key factor in stress assignment in verbs is the presence of a semantically opaque prefix. In the dataset on the stress patterns of verbs which will be presented in §3.4.2.1, if we exclude prefixed words and consider only verbs with no morphological structure at all, we find only four verbs which end with a diphthong, three of which have penultimate stress (*árgue*, *fóray*, *íssue*) and one which has final stress (*shàmpóo*). If we include word-final /ɪj/, we get an additional 23 verbs with penultimate stress (e.g. *cárry*, *húrry*, *márry*, *stúdy*, *wórry*). By contrast, the 31 prefixed disyllabic verbs that end in a diphthong all have final stress (e.g. *arráy*, *convéy*, *displáy*, *relý*,

¹⁷ As was seen above, available phonetic studies on that matter do not set diphthongs apart so any explanations of this kind may still be premature as long as we do not have robust evidence that diphthongs differ from other vowels in that regard. Moreover, breaking before liquids is a quite common process in the history of English which has affected various vowels, and not only monophthongs. Bourcier (1978: 155-164) discusses such cases and assumes that the insertion of what he calls a “glide” [ə] or [u] opened the way for vowel modifications and sometimes to the loss of the liquid. One illustration of this is the insertion of [u] before [t] in *talk* at the end of the 15th century, yielding [taułk].

¹⁸ No sources are provided for those pronunciations and it would be reassuring to have some evidence that this is how those words were adapted (and not, for example, simply by having the last consonant as a syllabic consonant, so that the underlying representations of those words would be |mɪjtɪ| and |tɛmplɪ|).

supply). Therefore, if prefixation is taken into consideration, then Polgárdi's counterargument has no empirical basis.

The elements reviewed in this section justify a change in the symbols that we use to transcribe standard British English. Therefore, in the rest of this document, I will be using the symbols in (1) to represent the vowels of SSBE. Moreover, I will use the symbol /ɪ/ instead of /r/. Most of the transcriptions that I will use come from Wells (2008) or Jones (2006) and will be adapted accordingly. Likewise, the examples that I will be taking from work that I have previously published will be adapted to those transcription standards. Transcriptions that are taken from other sources will be adapted to those standards when feasible, unless mentioned otherwise.

2.1.2. *Stress and accent*

There is little dissensus regarding the syllables that bear the main prominence in words, but the literature is less consensual on lower levels of prominence. As my dissertation topic was secondary stress, I had to seriously investigate what exactly that meant, and I often wondered about the source of the disagreements that I found in the literature: were different researchers disagreeing on the definition of stress, or were they actually describing different objects? I eventually came to assume that the second option was the correct one, and that there are two objects that have to do with prominence in English, '**stress**' and '**accent**'. A first version of this reflection appeared in my dissertation, in subsequent papers (Dabouis 2019, 2020b) and for a joint presentation with Marie Gabillet for the 2024 ALOES conference (Dabouis & Gabillet 2024). In this section, I mainly focus on the arguments developed in this presentation.

2.1.2.1. *Terminological confusion?*

In overviews of the terms "accent" and "stress" (Fox 2000: §3.1.1; Schane 2007; van der Hulst 2012, 2014), we can see how inconsistently the two terms are used. For some, "stress" refers to an abstract category with the potential for "accent", and "accent" is its concrete realization, while others use the two terms in the opposite way, "accent" being a potential "stress". There are also disagreements on the level at which those two objects are found, with some assuming that "stress" is a property of words while "accent" is a property of sentences. Overall, researchers would quite broadly agree on the strongest and the weakest syllables, but intermediate levels of prominence are where the controversy lies. To an extent, the difference is one between different traditions, with the American tradition being more 'generous' in subsidiary "stresses" than the British tradition, whose concern is, according to (Fox 2000: 127), "less with theoretically adequate formalization than with the identification of pedagogically useful categories". Examples that are assumed to have different "stress" patterns in those different traditions are shown in Table 2.

“American tradition” e.g. Pater (2000)	“British tradition” e.g. Wells (2008)
còndèmnátion	còndemnátion
Òctóber	Octóber
privátion	privátion
rétrogràde	rétrograde

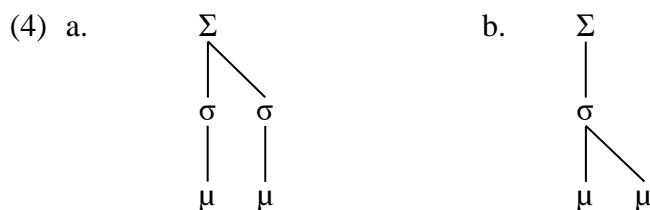
Table 2. Examples of differences in subsidiary prominence between the American and the British tradition

However, Fox (2000: 132) notes that differences in idiolects and differences in the aims pursued by the scholars of those different traditions (theoretical vs. pedagogical) cannot explain all the discrepancies found between them. Therefore, in the following sections, I will argue that those differences are not just terminological differences or different definitions of the same phenomenon, but rather that the prominence patterns represented in Table 2 refer to different phenomena, stress and accent. “Stress” will be assumed to be one kind of prominence, which is partially related to vowel quality, and has several segmental correlates which will be reviewed in §2.1.2.4. As will be seen in the coming sections, those correlates have often been captured using foot structure. As for “accent”, it will be assumed to be an additional level of prominence that may be assigned to a subset of stressed syllables. The issue of which stressed syllables may be accented will be covered in §2.1.2.5.

2.1.2.2. *Stress and feet*

Since early Metrical Phonology (Halle & Vergnaud 1978; Hayes 1981; Kiparsky 1979; Liberman & Prince 1977), and even more so since the advent of Prosodic Phonology (Nespor & Vogel 1986; Selkirk 1980b), prominence relations between syllables have been captured using feet, and so it seems useful to introduce this idea here as we will be needing it further on. In Prosodic Phonology¹⁹, the foot is one of several “prosodic constituents” in the “Prosodic Hierarchy”.

In that theory, stressed syllables are the head of a foot, and many phonological processes that refer to stress have been analysed in terms of foot structure. Common assumptions about feet in English are that they are trochaic (i.e. left-headed) and bimoraic (Bermúdez-Otero 2018b; Bermúdez-Otero & McMahon 2006; Pater 2000). This means that feet should either be made up of two light syllables (4a) or one heavy syllable (4b). Note that Σ represents a foot, σ a syllable and μ a mora.



¹⁹ This theory will be discussed in more detail in §3.2.1, especially regarding how it models the interface between morphosyntax and phonology.

As pointed out by Trevisan (2015: 9), the absence of use of metrical structures in the Guierrian School is a recurrent criticism “chiefly voiced by English-speaking phonologists”. He notes that this is a “valid objection in the case of iambic regression”, but in fact the criticism can be considerably broadened: prosodic structure has been used to capture generalizations about prominence and about vowel quantity but also for a number of segmental processes (to be discussed in §2.1.2.4). Those processes have not been investigated by Guierrian authors, which means that the scope of the models that they have developed may be very efficient for word-level prominence and vowel patterns, but they are limited to those phenomena. As feet have been used to unify a number of different phonological processes and are widely used by many phonologists, I contend that foot-based analyses should systematically be considered. I have done so in some of my work, and will revisit some of my previous studies using foot-based analyses in this document.

2.1.2.3. *The indirect relation to vowel quality*

The vowel system proposed by Lindsey and seen in §2.1.1 has a greater number of vowels which may be both stressed or unstressed (in the sense that they may or may not display the phonological behaviours detailed in the following sections) than the one that is adopted in reference pronunciation dictionaries. However, vowels (or syllabic consonants) of this kind are found in all systems, as illustrated in Table 3, in which we can find Wells’s (2008) system for British English, Hayes’s (1995) for American English and Szigetvári and Lindsey’s system for SSBE that was presented in previous sections.

	Wells (2008)	Hayes (1995)	Szigetvári / Lindsey
Always stressed / strong	i: e æ ɜ: ʌ əʊ u:	ey ε æ a ɔ ʌ U u	ɪ: ε(j) ε: a(j/w) ə: o: ɔ
	ɔ: ɒ ɑ: eə ɪə ʊə	ɪ oʊ before C	oʃ ɑ:
Variably stressed / strong	ɪ ʊ	ɾ ɹ²⁰ ɪ oʊ before V or # ɪ before ŋ yʊ ~ yə	ɪ(j) ɵ uʷ ə(w)
Always unstressed / weak	ɪ u ə	ə ɱ ɱ	ɪ ɱ ɱ

Table 3. Vowels and stress in different vocalic systems

In Wells’s system, there are five vowels which may be “weak”, two of which may also be “strong”, /ɪ/ and /ʊ/. In Hayes’s system, certain syllabic consonants may be stressed too, and FLEECE, GOAT, KIT and GOOSE may be unstressed in certain positions. In Szigetvári and Lindsey’s system, only non-low vowels and their corresponding diphthongs may be unstressed, and sonorants are assumed to be systematically unstressed.²¹ Thus, vowel quality plays a role, and the system may be said to be partially “quality-driven” (Kenstowicz 1997) if one assumes

²⁰ Hayes illustrates the possibility of stressing a syllabic /l/ with the word *pull*. Most other analysis assume that this word has a vowel and a consonant in its rime.

²¹ Szigetvári (to appear) leaves that possibility open for syllabic /l/ in rhotic accents for NURSE words.

that stress is ‘assigned’ to vowels, but quality alone is not always sufficient to tell stressed and unstressed syllables apart. Moreover, Hayes (1995) notes that stress is “parasitic” in that it involves phonetic resources that are used by other phonological phenomena such as pitch, intensity and duration, and that no clear phonetic correlate of stress can be found. Consequently, phonologists have looked for segmental cues that can help identify stressed syllables.

2.1.2.4. *The segmental diagnostics of stress*

A first series of processes that have been attributed to stress are those regulating the **allophones of voiceless stops**. While their distribution was initially attributed to syllable structure (Kahn 1976), since Kiparsky (1979), it has been attributed to foot structure: the aspirated allophones [p^h t^h k^h] are found foot-initially and lenited realizations (e.g. [ʔ r tʻ]) depending on the variety of English) are found in non-initial positions. The subsequent literature has shown that the distribution of /h/ and /ɹ/ in certain “broad” non-rhotic varieties of English showed a similar distribution (Davis & Cho 2003; J. Harris 2004, 2013), as shown in Table 4 and illustrated by the corresponding examples in (5).

		Word	
		Initial	Non-initial
Foot	Initial	(a) h t ^h ɹ	(b) h t ^h ɹ
	Non-initial	(c) h t ^h ɹ	(d) ɦ ʔ/r/tʻ ɻ

Table 4. The distribution of /h/, /ɹ/ and of the allophones of /t/ in different varieties of English (after J. Harris 2013)

- (5) a. [t^h]érrible b. dáven[p^h]ort c. [t^h]omáto d. á[r]om
 [h]ábit álco[h]ol [h]orízon véhicle
 [ɹ]ún Lé[ɹ]oy [ɹ]evíve húrricane

Crucially, the cases in (6b) are usually not assigned stress marks on their final syllable in British pronunciation dictionaries, but they pattern with the words in (5a), and with words with non-initial primary stress (e.g. a[t^h]áin, pro[h]íbit, de[ɹ]ánge). Using this kind of criterion has been argued to be a way to tease apart identical vowels when they are unstressed (6a) or stressed (6b) (examples from Szigetvári (2020)).

- (6) a. magnetic [mæg'neɹɪk] b. magnetism ['mægnət^hɪzəm]
 vanity ['væneɹɪj] manatee ['mæneɹt^hɪj]
 mosquito [mə'ski:row] ketone ['k^hɪjt^hown]

Note that different accounts have been proposed for the reason why there is aspiration word-initially in unstressed positions as in (5c). Some assume that aspiration can be found when voiceless stops are in the initial position of various prosodic constituents, including the phonological word (Itô & Mester 2009). So, we may assume representations such as (7a), as

syllable (10b). In terms of foot structure, this would be one of the rare cases in which a process refers to the right edge of the foot: here, the process occurs variably only if the syllable is in the final position of the foot.

- (10) a. fortunate [ˈfɔ:tʃənət] b. fortune [ˈfɔ:tʃʊwn] ~ [ˈfɔ:tʃən]
 botulin [ˈbɒtʃəlɪn] b. botulism [ˈbɒtʃʊwlɪzəm] ~ [ˈbɒtʃəlɪzəm]
 stimulus [ˈstɪmjələs] b. stimulate [ˈstɪmjʊweɪt] ~ [ˈstɪmjəleɪt]

Another process that is conditioned by stress, and has also been argued to be evidence for foot structure, is word-medial **syncope** (Hammond 1999, 2006; Szigetvári 2007, 2020). It has been argued that this process may occur before a stressless syllable, but not before a stressed syllable (in certain segmental configurations), as shown in (11). Hammond (1999: 165) argues that syncope may only occur if it improves footing, so that the resulting word is an optimal foot, and no syllable is left unparsed.²³ S. Davis (to appear) develops another view according to which syncope applies to a dactyl and generates a preferred trochaic sequence (and he details the segmental conditions in which it applies).

- (11) a. opera [ˈɒp(ə)ɪə] b. operatic [ˌɒpəˈɪatɪk]
 general [ˈdʒen(ə)ɪəl] b. generality [ˌdʒenəˈɪælətɪj]
 memory [ˈmɛm(ə)ɪj] b. memorize [ˈmɛməɪz]
 separate (adj.) [ˈsep(ə)ɪət] b. separate (v.) [ˈsepəɪt]
 barbarous [ˈbɑ:b(ə)ɪəs] b. barbarism [ˈbɑ:bəɪzəm]

Kiparsky (1979) also proposes that the foot is a domain of “**close contact between segments**” in the foot-medial position. Thus, he analyses several processes as having the foot as their domain of application. The first is the obligatory assimilation of /n/ to a following velar. That analysis is taken over by Bermúdez-Otero (2018a), who gives the examples in (12).

- (12) a. (Σ° **sing**) -ng- contained within Σ° → assimilation
 b. (Σ' (Σ° **cón**) **ga**) -ng- contained within Σ' → assimilation
 c. (ω (Σ° **cò**n)(Σ° **créte**)) -nc- split between feet → no assimilation²⁴

The other three are the mutual assimilation of [k] and [ɹ] (which he transcribes as [KR]) and illustrates with the pairs *crew*, *increase* (with mutual assimilation) and *back-rub*, *cock-roach* (without), the short variants of PRICE and MOUTH before a voiceless consonant (this is probably a form of pre-fortis clipping), and the devoicing of /l/ after a voiceless consonant (e.g. *ice-lip* [äjslɪp] vs. *eye-slip* [ajslɪp]). However, all the examples that are given to illustrate the absence of such close contact are compounds, and so the segments not only belong to different feet, but

²³ Péter Szigetvári (p.c.) argues that this explanation invoking improved footing cannot account for cases in which there cannot be an elision of the third syllable of an initial dactyl in *hullabaloo* /,hələbəˈlʊw/ or *methodological* /,mɛθədəˈlɒdʒɪkəl/. Indeed, that syllable does not belong to the previous foot and may be analysed as unparsed, excepted if one adopts the analysis in (14) below for precisely words with that configuration. In that case, the syllable is seen as foot-initial although unstressed, and that may be the reason why it is protected against syncope.

²⁴ Here to be understood as ‘non-obligatory assimilation’.

also phonological words (see e.g. Bermúdez-Otero 2011; Booij & Rubach 1984; Raffelsiefen 2007). Therefore, it is unclear that the difference between those pairs can be attributed to feet. In the case of pre-fortis clipping, the environment may be foot-bound as suggested by Kiparsky but in more intricate ways, as the relative strength of the two feet may be of importance (see Bermúdez-Otero 2004).

Finally, Szigetvári (to appear) mentions three processes that affect stressed and unstressed syllables differently: **compensatory lengthening** due to the loss of coda /ɪ/ (13a), **smoothing** of diphthongs followed by /ə/ (13b) and **syllabic consonant** formation (13c).

- | | | | | |
|------|----|--|-----|--|
| (13) | a. | defer /dɪ'fəː/
intern /ɪn'təːn/ ²⁵ | cp. | differ /'dɪfə/ *-/fəː/
pattern /'pʰatən/ *-/təːn/ |
| | b. | idea [aj'dijə] ~ [aj'dɪː]
secure [sɪ'kjʊwə] ~ [sɪ'kjɔː] ²⁶ | cp. | India ['ɪndɪjə] *[-dɪː]
jaguar ['dʒɑgjuwə] *[-gjoː] |
| | c. | tunnel ['tʰənəl] ~ ['tʰənɫ]
caramel ['kʰɑɾəməl] ~ ['kʰɑɾəmɫ] | cp. | annul [ə'nəl] *[-'nɫ]
philomel ['fɪləmɛɫ] *[-mɫ] |

I do not know of foot-based analyses of those processes, but we could analyse compensatory lengthening as occurring only if the syllable is a foot head, smoothing as affecting only vowel sequences in which the first one is stressed (and it could be said that it improves parsing by incorporating the following /ə/ into a minimal foot), and syllabic consonants may be analysed as being banned from foot-head position.

Many of the processes reviewed in this section show that **certain post-tonic vowels behave just as primary stressed vowels**. Then, if stress is what is behind those processes, they must also be analysed as stressed. This suggests that, contrary to what can be found in the transcriptions of pronunciation dictionaries, post-tonic secondary stress is actually quite common. Words with a primary stress and a post-tonic secondary stress will be referred to as 'left-prominent' while those with the reverse relative prominence will be referred to as 'right-prominent'.

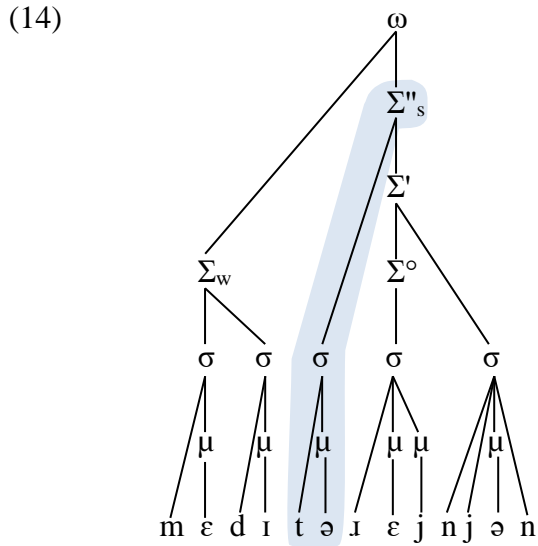
In this section, we have seen how many of those processes have been or can be analysed using foot structure. However, there is one case in which the left edge of a foot has been argued not to correspond to a stressed syllable: pretonic syllables preceded by a bimoraic trochee.²⁷ Those have been assumed to adjoin to the following foot, thus behaving as foot-initial even

²⁵ Szigetvári only uses examples in which the vowel has primary stress, which is usually accented, and so Jean-Michel Fournier (p.c.) argues that this could be attributed to accent. However, we do find cases of post-tonic /əː/ (e.g. *intern*_N /'ɪntəːn/, *taciturn* /'tasɪtəːn/). However, it is unclear whether the contrast in (13a) is really to be attributed to compensatory lengthening rather than to a historical process of reduction. Moreover, the lengthening in stressed position does not seem to have occurred to compensate for the loss of /ɪ/ but actually preceded it (Wells 1982: 201).

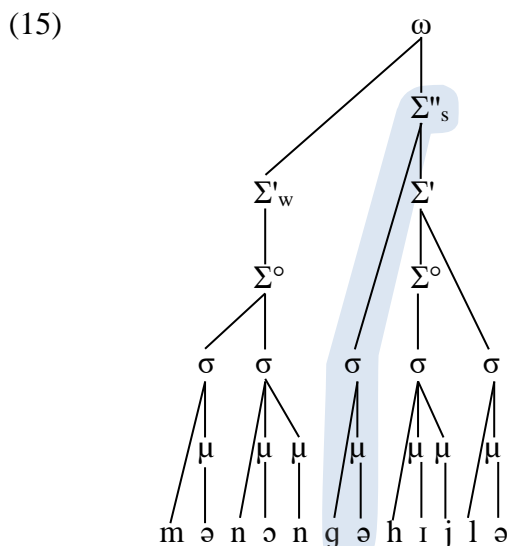
²⁶ Here too, only examples with primary stress are given, and so we could assume that this is restricted to primary stress. However, this seems incorrect as certain post-tonic vowels may undergo smoothing (e.g. *empire* ['em,pʰɪjə] ~ ['em,pʰɑː]; *safflower* ['sa,flawə] ~ ['sa,flɑː]). Wells (1982: 239) notes that this is quite common for /aw/ and /aj/.

²⁷ Actually, the number of pretonic syllables might be relevant, as those processes do not appear to affect the second syllable of words in $\bar{\sigma}\sigma\acute{\sigma}$.

though they are unstressed (Bermúdez-Otero 2018a; S. Davis & Cho 2003; J. Harris 2013; Jensen 2000). This is diagnosed through the absence of flapping and presence of aspiration for voiceless stops in American English in those positions (e.g. *Mèdi*[t^h]*erránean*) or the presence of word-internal /h/ before a stressless /ə/ (e.g. *Tàrahhumára*). In those words, the assumed foot structure is as represented in (14) for *Mediterranean*, with the third syllable adjoined to the following foot, making it foot-initial.



Independent evidence for foot structures such as the one in (14) has been argued to come from non-obligatory assimilation of the /n/ in *Monongahela* (Bermúdez-Otero 2015) which suggests, following Kiparsky (1979), that /n/ and /g/ belong to different feet. Therefore, the foot structure of that word can be assumed to be as shown in (15).



Let us conclude this section with Table 5, which summarizes all the processes that have been reviewed here, and how they may be described using either stress or foot structure.

STRONG AND WEAK SYLLABLES

Process	Stress-based generalization	Foot-based generalization
Aspiration of /p, t, k/ Presence of /h/ Presence of /ɪ/ in broad non-rhotic varieties	In the onset of a stressed syllable, word-initially or in the onset of pretonic syllable which is preceded by a minimal foot	Foot-initially
Epenthesis between a nasal and a voiceless fricative Assimilation of /n/ before a velar	Blocked if the fricative is in the onset of a stressed syllable Optional if the velar is the onset of a stressed syllable	Within a foot
Yod-dropping Compensatory lengthening	In a stressed syllable	Within a minimal foot
Loss of /w/ in GOOSE	Variable if not followed by an unstressed syllable	Variable in the last dependent syllable of a foot
Syncope Smoothing	Blocked before a stressed syllable If the first vowel is stressed	If syncope improves foot structure If the first vowel belongs to the head of a foot
Syllabic consonant formation	Blocked in stressed syllables (except /ɪ/ in GA)	Prohibited for a foot-head

Table 5. The segmental diagnostics of stress proposed in the literature for English

Let us also mention that all of those claims should be taken with caution. While the evidence seems to converge on the fact that there are more secondary stresses, notably post-tonic ones, than indicated in pronunciation dictionaries, this evidence is very often quite scarce, and very often relies on intuition or anecdotal evidence. Thus, there is a great need for a closer investigation of the segmental correlates of stress.

2.1.2.5. *The interaction of stress and accent*

Some of the recent literature on those issues assumes that accents anchor onto a subset of stressed vowels (Gussenhoven 2004, 2011; Szigetvári 2020). The question is then to know which ones. Are there differences between left-prominent words and right-prominent words?

Recent phonetic studies that have looked at right- or left-prominent noun-noun compounds (Kunter 2011) and derivatives (Plag & Kunter 2007; Plag *et al.* 2011) and have found that words with different prominence patterns behave differently. For those two types of words, those studies have found that left-prominent words only receive a single pitch accent while right-prominent words receive two. Kunter's study on compounds reports no significant differences in pitch and intensity between the two constituents of a compound in right-prominent compounds (e.g. *Pàrk Avenue*) using corpus data, but finds higher pitch and intensity in the first element of a left-prominent compound (e.g. *windshield*). Those results are consistent with previous acoustic studies of prominence in compounds (Farnetani *et al.* 1988; Plag 2006) and with the autosegmental-metrical model of intonational phonology (Gussenhoven 2004). Similarly, Plag *et al.* (2011) compare primary stress and secondary stress in right-prominent (e.g. *viòlâtion, pùblishée*) and left-prominent derivatives (e.g. *rándomìze, àctivàte*) in accented and unaccented position in a production experiment. They find strong differences between the

two types of words in F_0 , intensity and spectral balance between the two stressed syllables in both types of words, and in both accented and unaccented positions, although those differences are very marked for left-prominent words, and they are much smaller for F_0 and intensity in unaccented positions. They comment that

This implies that the notion of secondary stress is to some extent problematic. While the primary stress syllable and the secondary stress syllable are both strong syllables irrespective of their respective positions within the word, the acoustics as well as the phonology (in terms of accentuation) of the two secondary stress syllables in right-prominent vs. left-prominent words are quite different from each other. (Plag *et al.*, 2011)

Moreover, they analyse those results as showing that right-prominent words receive two pitch accents while left-prominent words only receive one if placed in accented positions and that, in unaccented positions, there is no clear phonetic difference between primary and secondary stress.

Those observations could lead one to believe that post-tonic secondary stresses never affect accents, but there is one intonational contour in which the placement of pitch-accents depends on the position of secondary stresses: the chanted vocative (Gussenhoven 2004: 57; Hayes 1995: §2.3.3; Liberman 1975: 30). In that contour, the default placement of the M^* is the last syllable of a word (16b, c), unless the penult carries secondary stress (16a).²⁸

- (16) a. Poindexter! b. Annabel! c. Pamela!
- /'pɔɪn,dekstə/ /'anə,bel/ /'pamələ/
- | / / | / |
- H* M* H* M* H* M*

The placement of accents on pretonic syllables that carry secondary stress is less documented. Hayes claims that the “surprise-redundancy” contour is a good diagnostic of stress, and that in this contour an L^* is associated with the strongest stress preceding primary stress, as in *collàborátion* or *clàssificátion*. The issue is more problematic with words with multiple pretonic secondary stresses, for which speakers may associate the L^* to one of the first two syllables if they both carry secondary stress, as in the examples in (17).

- (17) a. sensationality! b. Constantinople!
- /senseɪʃənələtɪj/ /kɒnstəntɪnəwpəl/
- | | | | | |
- L* or L* H* L* or L* H*

Once again, this suggests that words may have several accents (*contra* van der Hulst 2012, 2014a, b), although accents may be deleted. This is the approach taken by Gussenhoven (1991, 2011), who assumes that accents are assigned lexically and may be deleted post-lexically. In this approach, the “Rhythm Rule” can be seen as a post-lexical accent deletion rule (see also

²⁸ In that approach to intonation, H refers to a high tone, M to a mid tone and L to a low tone. * indicates that the tone is associated to a prominent syllable.

Bolinger 1986; Horne 1990), as shown in the examples in (18), in which the second lexical accent of *thirteen* and *Japanese* may be deleted to avoid an accent clash.

(18)	a.	thirteen men ,θə:'tɪjn 'mɛn	b.	Japanese chairs ,dʒapə'nɪz 'tʃɛ:z
	Lexical accents	* * *		* * *
	Post-lexical accents	* *		* *

This means that words may have pretonic syllables with secondary stress that receive accents that are adjacent to primary stress, at least lexically, and that such accents may or may not be realized. Gussenhoven (1991, 2011) assumes that at the word-level, pretonic secondary stresses get accented when they immediately precede primary stress if the word is disyllabic (e.g. *sardine* /,sɑ:'dɪjn/) or if it is prefixed (e.g. *unmodest* /,ən'mɒdəst/), but not in other configurations (e.g. *October* /,ɔk'təwə/).

The view developed by Gussenhoven consists in assuming that stress comes first, and then that accents are assigned to certain stressed syllables, and that some of those accents may be deleted post-lexically. Under that assumption, the phonetic studies reviewed above show that post-tonic secondary stresses normally do not get accented (except in the chanted vocative) while pretonic stresses usually do, unless they are deleted post-lexically. If one assumes that ‘stress-then-accent’ view, then it is necessary for phonological theory to represent the difference between primary and secondary stress to account for the different accentuation patterns between right-prominent and left-prominent words, and so stress cannot be binary (*contra* Szigetvári 2017). According to van der Hulst (2014a), the ‘stress-then-accent’ view is the approach of “standard metrical theory”, but he adopts an “accent first” approach, in which an accent entails a stress, but is only one of the possible sources of stress, along with edge-marking, rhythm, cyclicity, weight and lexical marking. The approach taken in the Guierrian School can actually be said to be closer to this approach, where accents are assumed to come first (but there can be more than one accent per word, unlike in van der Hulst’s approach), and then “vowel reduction” (which some would call “destressing”) may apply. It is also possible that the whole question of knowing which ‘comes first’ does not make sense and that both are always present lexically and that accents may not get realized. The evidence from the phonetic studies reported above may actually support the view that stress is binary, as primary and secondary stressed vowels are indistinguishable from one another in unaccented contexts. What would distinguish left-prominent from right-prominent words would then be the presence or absence of an accent on the rightmost stressed syllable.

Another more fine-grained view is put forward by Fox (2000: 145), who argues that it is important to take into account “the difference between *actual* and *potential* accents” (emphases in the original). He argues that there may be variation in the number of accents that will be realized for a given word or expression depending on its context, the speech style or speed of utterance, and that this may create “the impression of different degrees of accent or stress, since such ‘degrees’ reflect a hierarchy not of accents themselves but of potentiality for

accent”. To illustrate this, he uses the example of the expression *elevator operator*, which may have between one and four accents, as illustrated in (19).²⁹

- (19) *the 'elevator operator's 'car*
 '*elevator 'operator*
 '*ele 'vator 'ope 'rator*

Thus, the number of accents in a word may depend on the context in which the word is uttered, but it is to be noted that there is still some hierarchy, with stronger stresses being more likely to be accented.

Fox also argues that there is a first level of accentuation which has to do with rhythm, and roughly corresponds to stress and foot structure as we have described them in the preceding sections, and a second level of accentuation that has to do with intonation, which does not depend on rhythm. One indication supporting the view that stress patterns are indeed conditioned by rhythmic considerations in my own research has been the observation, in my dissertation and in a subsequent paper (Dabouis 2020b), that pretonic strings of more than two unstressed syllables were really hard to come by. Using dictionary data, I found only 84 words which have at least four syllables before the primary stress, and only five of those words are not derived from another word. Among those, only one word may have three fully unstressed syllables in a row (*taramasalata* /*ˌtɑːrəməsəˈlɑːtə/* ~ /*təˈrɑː-/*) while the others have a medial stressed vowel which is not necessarily marked as such in the dictionary (e.g. *ipecacuanha* /*ˌɪpəˈkækjuˈwɑːnə/*) or /*ɪ/*, which may be analysed in the same way (e.g. *prestidigitation* /*ˌpreʃtɪˈdɪdʒɪˈteɪʃən/*). In morphologically complex words, we find cases which are like the last two categories found in non-derived words (e.g. *justifiability* /*ˌdʒʌstəˈfæjəˈbɪlətɪj/*, *desertification* /*ˌdezəˈtɪfɪˈkeɪʃən/*), and cases in which the potential string of unstressed syllables may always be reduced to only two unstressed syllables in a row (e.g. *criminalization* /*ˌkrɪmɪn(ə)lɪˈzeɪʃən/*, *sexualization* /*ˌseks(ʊ)wəlɪˈzeɪʃən/*). Apart from *taramasalata*, no word has three unstressed syllables in a row, and that word may possibly be analysed as a kind of ‘crypto-compound’ (as suggested by Bermúdez-Otero p.c.), which would mean that the two putative constituents in that word would have phonological words of their own so that no phonological word contains a string of three unstressed syllables: (tɑːrəmə)_ω(sələːtə)_ω. Similarly, Dahak (2011) reports that “full” vowels are more often found in the post-tonic position if the position considered is not immediately adjacent to primary stress (/ˈσσσσ/ vs. /ˈσσσ/, /ˈσσσσ/, /ˈσσ/, /-ˈσσ/). However, Fox’s position on a second level of accentuation that would not be rhythmical is challenged by cases such as those in (18), in which the loss of an accent seems to be driven by the avoidance of adjacent accents, which may be seen as an effect of rhythm.

Before concluding this section, let us say a word on pronunciation dictionaries in the British tradition such as Wells (2008) or Roach (2009), which have been used extensively in studies conducted in the Guierrian School, and as references for French competitive exams (CAPES and agrégation), thus strongly impacting how English phonetics and phonology are

²⁹ Another example of the loss of a strong accent might be the observation that semantically transparent prefixes, which are usually assumed to have their own phonological word, and so to be quite prominent, may get entirely reduced in speech (e.g. *rebuild* [ɪəˈbɪld], *unfortunately* [(ə)nˈfoːtʃənli]), as shown by Videau (2013), Videau & Hanote (2015) and Hay (2007).

taught in France. Therefore, it is worth asking the question of what exactly is represented by the “stress marks” , and ' in those dictionaries. In the light of what we have seen, it seems they mostly mark accents (see Abercrombie (1976) and Schane (2007) for similar views) or, in line with Fox’s view on potential accents, commonly accented syllables. However, they tend to mark certain post-tonic syllables as bearing a “secondary stress”, notably the second constituent of a left-prominent compound if its second constituent has at least two syllables (e.g. *countrywoman* /'kʌntri ,wʊm ən/ cp. *countryside* /'kʌntri i saɪd/ in Wells (2008)) and words containing the suffix *-ism* (e.g. *activism* /'æktɪv ,ɪz əm/). As was seen above, those syllables will usually not bear an accent (except in the chanted vocative), but they can certainly be analysed as stressed using the criteria seen in §2.1.2.4. This use of post-tonic stress marks is inconsistent because they do not mark the last syllables of words such as *absolute* or *conjugate* as stressed, and it is unclear why they should not be marked in the same way as compounds and words in *-ism*.

As was seen previously on the vowel system or on the segmental processes associated with stress, the evidence on the interaction between stress and accent is quite scarce and would require much more detailed investigation. However, what is available seems to converge on the fact that those are indeed two different objects. We have seen that stress is partially related to vowel quality, as certain vowels are systematically associated with the segmental processes associated with stress, others are only variably so, and yet others never are (in the system adopted here, the latter category only concerns syllabic consonants). We have also seen that what is usually described as “stress” in British pronunciation dictionaries is in fact (commonly) accented syllables and some cases of post-tonic secondary stress. Because of this bias induced by the type of source I have been using in most of my work, the studies that I will be reporting mainly focus on the distribution of accents (or at least, accents that are commonly realised), which corresponds mainly to primary stress and (some) pretonic secondary stresses. In the next section, I turn to the question of the relationship between stress and segmental structure, which was partly covered here when I discussed the relationship between stress and vowel quality.

2.1.3. *Stress and segmental structure*

One of the positions that has been central to the Guierrian School is its critical view regarding the claim made in the generative tradition that segmental structure is a key determinant of stress (see e.g. J.-M. Fournier 2010a; Guierre 1983). The most common claim is that primary stress is conditioned by the structure of the penult in nouns and that of the final syllable in verbs (see §3.4.1 for more detail). While the role of what is now thought of as syllable weight was central in *SPE*’s stress system, it has been considerably nuanced in subsequent work. For example, sonorants and /s/ have been analysed as having variable weight (Burzio 1994; Giegerich 1999; Halle & Vergnaud 1987; Selkirk 1984), and the effects of weight are now more seen as probabilistic effects (Domahs *et al.* 2014; Moore-Cantwell 2016, 2020 and see §3.4.2.2). In nouns, it has also been shown by Pater (1994) that, while nouns with heavy penults quite reliably have penultimate primary stress, those with light penults have a more variable behaviour, split mainly between penultimate and antepenultimate stress. However, despite those nuances, syllable weight is still assumed to be one of the key determinants of the position of stress.

A question to be raised is which parts of segmental structure can be reliably shown to have an effect on stress (or accent). As mentioned above, segmental structure is mainly assumed to affect stress through syllable structure (through syllable weight), and only in certain positions. The part of syllable structure that is generally claimed to affect stress is the rime (nucleus and coda) but not the onset. Recent work has challenged this claim, as there have been a series of studies reporting effects of the number of onset consonants in disyllables on primary stress (Kelly 2004; Ryan 2014; Treiman *et al.* 2020): more consonants in the initial onset were found to favour initial primary stress. Ryan also reports effects of the onset of the second syllable that are comparable: more onset consonants are associated with more stress on that syllable. However, it has not been tested beyond those particular configurations. As to positions in which syllable weight affects the position of stress, several recent experimental studies using pseudowords have reported that weight (implemented as closed syllables vs. open syllables) played a role in all positions, contrary to widespread assumptions (Domahs *et al.* 2014; Garcia & Goad to appear). As was mentioned above, even in the positions in which primary stress is thought to be affected by the presence of coda consonants, not all consonants behave alike, and some of the most common³⁰ (potential) coda consonants (sonorants and /s/³¹) have been claimed to have variable weight. This means that the association between their presence in relevant positions and primary stress is not systematic, and sometimes close to chance. In a large dictionary-based study, Moore-Cantwell (2020) treats those ambiguous categories apart and finds that the rate of penultimate stress falls in between light and heavy syllables, close to 50%. As will be seen in §2.2.2, the place of the coda consonant has also been claimed to affect stress, through a distinction between coronal and non-coronal consonants. The study that will be discussed in §3.4.2.1 shows that there is strong underdetermination in the predictors of primary stress in verbs because of the strong correlation between predictors, especially weight and opaque prefixation. Finally, as was seen in the previous section, certain vowels are always stressed, although they are not always accented.

Let us explore the issue of the relationship between stress and vowels a bit further. In the penultimate position, there is a strong correlation between the presence of long vowels (or diphthongs – but to simplify the discussion here, let us simply talk about long vowels) and primary stress. The usual assumption in the generative tradition is that stress is assigned to penultimate syllables that contain long vowels. Thus, in a word like *horizon* /hə'ɹɔjzən/, the underlying representation is assumed to be [həɹɔjzən], and it is the presence of [ɔj] in the penult that would explain the fact that primary stress is placed on that syllable. However, from the perspective of the learner, words are encountered orally with both pieces of information at the same time, and assuming that stress is derived from vowels or the reverse means that we are assuming that only one of those two things is kept in a lexical (or underlying) representation. I am not aware of any arguments that would explain why that would be. However, there is a lot of literature on the fact that most words are stored along with their prosodic specifications (e.g.

³⁰ In Dabouis (2016c), using a dataset of 1865 suffixal derivatives, I found that 688/998 (69%) of internal codas were either a sonorant or /s/. Moore-Cantwell's (2020) results show that, in words of at least three syllables, the proportion of such ambiguous cases is predominant in the penultimate syllable: 913/1182 (77%). Both counts show that those ambiguous cases constitute a large majority of (potentially) closed penults.

³¹ I say 'potential' because the syllabification of /sC/ clusters is notoriously controversial.

see the discussion of nonanalytic listing in Stratal Phonology in §3.2.2).³² We can thus hypothesize that learners store both vowels and stress at the same time, and that they construct a grammatical system that uses generalizations capturing the relationship between them, but also other parts of linguistic structure (e.g. morphology or semantics). However, there is another way that unknown words can be encountered and where neither vowel length nor stress are available: through orthography. Thus, the early generalizations that the learner has extracted may be complemented by graphophonological generalizations capturing the regularities found between orthographic and phonological representations (see Ch. 5) as they become literate. The Guierrian tradition has explored in some detail such graphophonological generalizations, and so the relationship between vowel length and stress is seen in the opposite direction as that which we have just seen: it is assumed that penultimate vowels that are stressed should be long. Using the same example of *horizon*, knowing that the penult is stressed predicts that the vowel will be long. This approach can allow us to establish which generalizations are useful to assign stress to an unknown word based on segmental structure in an orthographically-informed phonological system:

- The presence of certain consonantal strings may be associated with the presence of stress in certain positions. This includes orthographic clusters which are mapped onto a single phonological consonant (e.g. *vanilla*, *grammar*, *dilemma*, *Kentucky*). This is particularly true for primary stress in non-native parts of the vocabulary (see Ch. 5) and in some of the studies that I co-authored, it was found that this was also true for pretonic secondary stress placement (see §2.2.2 and §2.3). It remains to be established whether consonants that are syllabified as onsets can be shown to be generally correlated with the position of stress beyond disyllabic words;
- Vowel digraphs generally represent diphthongs or long vowels, and so they are often stressed. There are several environments in which they are reliable cues to the position of primary stress (e.g. when <oo> or <ee> are in the final position, or when a digraph is found in the penultimate syllable of a neoclassical compound, e.g. *thérapeú*tist, *dínosá*urus). They are very often realized as stressed in syllables which do not bear primary stress, especially the initial pretonic position (e.g. *authentic* /,o:'θɛntɪk/, *routine* /,rʊw'ti:n/);
- There are several suffixes which were mostly neutral which have progressively evolved so that the words that contain them have primary stress close to the right edge of the word, and this may sometimes be attributed to the presence of long vowels in the final syllable of their base. For example, *-able* derivatives increasingly have primary stress on the presuffixal syllable if it contains a long vowel (e.g. *jústifiable* → *jùstifiable*). I have also confirmed existing claims of this behaviour regarding the adjectival suffix *-al* (Dabouis 2016a). As will be seen in §3.5, this can be analysed as a kind of prominence shift. In such cases, there is no issue with the claim that vowel length affects primary stress (and not the reverse) because vowel length may be assumed to be known in the base when the stress patterns of such derivatives are computed.

³² In the case of stress-shifting suffixes like *-ic*, *-ion* or *-ity*, there are studies showing that the acquisition of the stress patterns associated to those suffixes is sensitive to lexical frequency, which can be taken to suggest that the words that contain them are lexically stored (Jarmulowicz 2006; Jarmulowicz *et al.* 2008).

In this section, we have seen that segmental structure, and especially rime weight, is still thought to be one of the central predictors of stress in English. Therefore, in the studies that I have conducted or taken part in which will be reported in the following sections, it is quite logically that it will figure among the tested predictors. We have also seen that the empirical evidence and theoretical arguments are more robust regarding the effect of closed syllables on stress than that of long vowels, although questions remain regarding certain consonants which may close syllables but whose weight has been argued to be variable. Finally, we saw that, in the context of an orthographically-informed model, we can still attribute certain stress patterns to the presence of long vowels and have generalizations for both stress and vowels that may be used generatively in reading new words.

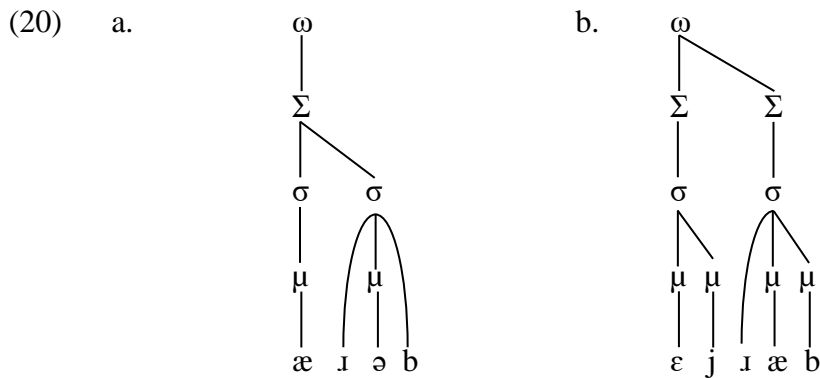
2.2. Vowel reduction

Since at least *SPE*, the process by which certain vowels weaken when found in weak syllables (e.g. *atom* /'atəm/ cp. *atomic* /ə'tɒmɪk/) has been referred to as “vowel reduction”. Based on what we saw in the previous section, this can be seen as a kind of destressing, or as a process that goes along with destressing. However, as most of the studies that I have conducted use the term “vowel reduction”, and in line with much of the literature which does too, I will keep using that terminology here. The terms “full vowel” or “reduced vowel” could very well be substituted by “(secondary) stressed vowel” and “stressless vowel” in this section. The studies discussed in the following sections focus on positions in which there is usually no accent: immediately pretonic vowels and post-tonic vowels. Note that there is controversy on the status of vowel reduction: is it really a synchronic phonological process or is it a historical relic, as suggested by Szigetvári (2018, 2020)? The view that I will adopt here is that vowel reduction is certainly not a fully regular, systematic phonological process. Rather, it is partly lexically conditioned and sustains lexical exceptions, and it is therefore probably best analysed as some form of lexical redundancy rule (Jackendoff 1975). This means that it mostly captures generalizations about forms stored in the mental lexicon, but may occasionally be used generatively. The generalizations regarding the distribution of unstressed vowels must also be a part of the ability of literate language users as English orthography does not indicate whether a vowel is full/stressed or reduced/unstressed. I will mention three studies on vowel reduction that I have taken part to, in chronological order. First, the study of the so-called “Arab Rule” (§2.2.1), then the study of pretonic vowels (§2.2.2) and finally the study of vowel reduction in French loanwords (§2.2.3). The aim of those studies has mostly been to put to the test several claims found in the literature using dictionary data, not to formalize the process (except for the Arab Rule).

2.2.1. First steps: *The Arab Rule*

As mentioned above, the framework in which I was trained, the Guierrian School, does not usually use feet in its analyses (although some metrical analyses can be found in Castanier (2016)). This led me to read extensively about this theoretical object and its uses. In an exchange on the issue, Ricardo Bermúdez-Otero mentioned the case of the so-called “Arab Rule” as

evidence for the moraic trochee being the default foot type in English. The name of the “Arab Rule” refers to two American idiolectal pronunciations of the word *Arab*: /'æɹəb/ and /'ɛjɹəb/ and states that vowels found in syllables closed by non-coronal obstruents reduce less than vowels found in syllables closed by coronal obstruents, although reduction is the norm if the first syllable is light, vowels normally reduce (Fidelholtz 1966; Hayes 1982; Pater 1995, 2000; Ross 1972). Thus, the pronunciations of *Arab*, which vary in the weight of the first syllable at the same time as the reduction of the second syllable, illustrate this. Using a foot-based analysis, those two pronunciations correspond to the structures in (20).³³ The key component of the analysis is that, if the first syllable is heavy, it constitutes a well-formed bimoraic trochee and so the second syllable forms one as well, while if it is light, the two syllables are part of the same disyllabic foot.



Thus, as existing studies did not quantify the phenomenon, the first thing I did was to manually collect all disyllabic words in Wells (2008) with first-syllable primary stress, a single non-coronal coda in their second syllable and no identifiable morphological structure. As vowel reduction has been noted to be influenced by word frequency (Bell *et al.* 2009; Clopper & Turnbull 2018; Fidelholtz 1975), I also collected frequencies for those words in SUBTLEX-UK (Van Heuven *et al.* 2014). Words were coded as having a first syllable that is light (e.g. *stomach* /'stɒmək/, *syrup* /'sɪɹəp/) or heavy (e.g. *Joseph* /'dʒəʊwsɪf/, *epoch* /'ɪj.pək/) and the reduction of the second vowel was determined using the transcriptions from Wells (2008): vowels were treated as reduced if they were transcribed as /ə/ or if they were an orthographic <e> transcribed as /ɪ/ or /ə/.³⁴ Based on an ordinal logistic regression analysis, the results confirmed a significant effect of the weight of the first syllable as a predictor of the nature of the vowel of the second syllable. This can be shown in Figure 2.

³³ Although I am not aware of any published foot-based analysis about this particular phenomenon word-finally, I assume that for it to be captured using moraic trochees, one needs to assume that the final /b/ does not project a mora in (20a) but does in (20b), probably to satisfy foot minimality. Another option might be to assume that the unstressed /ə/ is moraic (Hammond 1999: 206-207) but that /b/ does project a mora.

³⁴ The full dataset is available on OSF: <https://osf.io/bp27r/>.

STRONG AND WEAK SYLLABLES

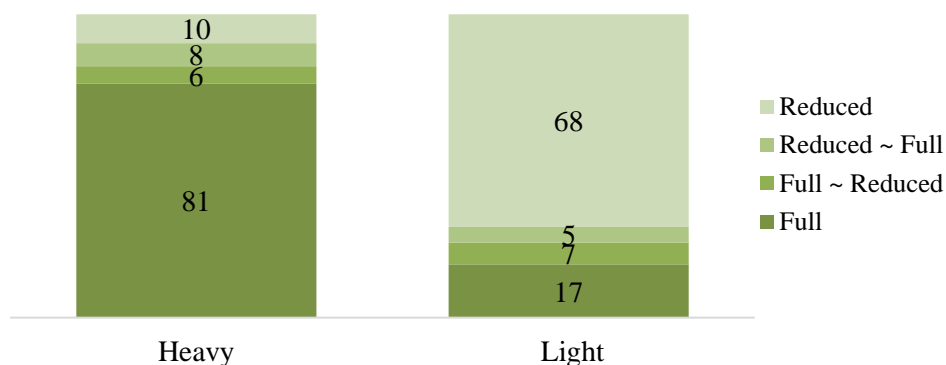


Figure 2. Nature of the vowel of the second syllable depending on the weight of the first syllable in disyllabic words in the context of the “Arab Rule”

Those results clearly confirm the interaction between the vocalic behaviour of the two syllables, although I did not test if such an interaction is indeed restricted to words in which the second syllable is closed by a non-coronal obstruent. I shared those results with colleagues working in the framework of Government Phonology, Nicola Lampitelli and Guillaume Enguehard, and we published them in a paper in 2020 with Jean-Michel Fournier (Dabouis *et al.* 2020). In that paper, we also develop an analysis of the process within Government Phonology (see §3.2.4), a theoretical framework that does not have feet and uses lateral relations to capture what Prosodic Phonology can capture using prosodic constituents (see §3.2.4).

In this study, we focused on the Arab Rule as it is described in the literature, with its two determinants: the weight of the preceding syllable and the presence of a non-coronal coda. However, it would be interesting to establish whether or not non-coronality is indeed necessary: are there effects of the weight of the preceding syllable with coronal codas too?

Interestingly, the Government Phonology analysis that we developed makes the prediction that closed initial pretonic syllables should not be reduced. During this study of the Arab Rule, I had started reading extensively about the determinants of vowel reduction that have been proposed in the literature, and this prediction to be tested led Jean-Michel Fournier and I to pursue our investigations on the distribution of reduced vowels further.

2.2.2. *The study of vowel reduction in pretonic syllables*

2.2.2.1. *Background*

After those initial investigations on the Arab Rule, we set out to conduct a more extensive study of vowel reduction. Our focus turned to pretonic syllables because they are the most widely discussed, and they are the positions in which morphological effects can be seen (notably the difference in the famous pair *cond/ε ~ ə/nsation ~ comp/ə/nsation*). Our results were first presented in several conferences in 2019 (MFM and PAC) and are discussed in more detail in a paper that has been accepted for publication *Phonology* (Dabouis & J.-M. Fournier to appear).

An important part of the study was first to review the literature on the matter. In doing so, one can discover a wide variety of claims that have been made regarding the factors which

may affect vowel reduction. In our accepted manuscript, those are reviewed in some detail, and they are here summarized in Table 6.

		Favours reduction	Disfavours reduction
Syllable structure		Open	Closed
Nature of the coda	Place	Coronal	Non-coronal
	Manner	Sonorant	Obstruent
Arab Rule		Preceding light syllable	Preceding heavy syllable
Position		Non-initial	Initial
Vowel features	Prevocally	Lax	Tense
	Word-finally	Low	Non-low
Frequency		High frequency	Low frequency
Foreignness		Native	Foreign
Orthography		Monograph	Digraph
Morphology	Stress-shifted derivative	Non-derived / no shift	Syllable has stress in base
	Status of initial syllable	Opaque prefix	Part of the root
	Status of final syllable	Non-unique syllable of the root	Unique syllable of the root

Table 6. The factors claimed to affect vowel reduction

What reviewing the literature also showed was the scarcity of empirical investigations on the factors listed in Table 6. We found three studies which sought to study some of those factors quantitatively (Dahak 2011; Hammond 2003; Tokar 2019; Zhang 2021) but all of them suffer from limitations such as focusing on specific classes of words, studying only a restricted set of predictors, or not using multivariate analyses. Therefore, our aim was to fill that gap in the literature by providing a large study of those factors.

2.2.2.2. Methodology

We decided to use a dataset of 4,494 words from pronunciation dictionaries as a first step in a multifactorial investigation of the phenomenon. This has the advantages of restricting the investigation to a relatively uniform idiolect (although it is somehow artificial) and gives us access to a large number of different words. As to the reliability of this kind of data for the study of vowel reduction, I conducted a mini-study in which I compared the pronunciations given in the dictionaries to oral data from YouGlish (<https://youglish.com/>, consulted between January 12 and March 9 2021) for 200 randomly selected words, focusing on vowels in different positions, in different syllabic configurations, with different orthographic vowels and in different kinds of words (non-derived words and stress-shifted derivatives). I collected a maximum of ten tokens per word and classified the vowel of interest as full or reduced by ear. It was found that the transcriptions of the dictionary, and especially the main pronunciations,

broadly align with the oral data although – quite expectedly – there is more variability in the oral data than in the dictionary data.

We chose to focus on two positions: the initial pretonic position (first syllable followed by a stressed syllable, e.g. *arrive*, *dexterity*, *heretical*) and the intertonic position (second syllable followed by the syllable with primary stress and preceded by a secondary stressed syllable, e.g. *notoriety pioner*, *relaxation*). We considered only three categories of words: what we called “non-derived” words (i.e. monomorphemic or bound root + suffix; e.g. *acacia*, *tarantula*, *ambition*, *sporadic*), prefixed words with a monosyllabic prefix which may be more or less opaque (e.g. *believe*, *collect*, *reactivate*, *unaltered*) and stress-shifted derivatives (i.e. in which primary stress is on a different syllable than the base of the word, and the syllable of interest has primary stress in the base; e.g. *vital* ♦ *vitality*; *inform* ♦ *information*). In this section, I focus on non-derived words. The results for prefixed words will be discussed in §3.3.1.1 and those for stress-shifted derivatives will be seen in §3.7.2.3. The number of words for each subset is shown in Table 7.

	Initial pretonic	Intertonic
1. Non-derived	1,234	474
2. Prefixed	1,997	-
3. Stress-shifted derivatives	590	199

Table 7. Word counts in the different datasets of the study on vowel reduction

The source of the data was different for the two positions. The initial pretonic dataset was initially designed as a follow-up to the Arab Rule study, and so it was extracted automatically from Jones (2006) using the Laboratoire Ligérien de Linguistique’s Dictionary Database and cleaned up and annotated manually. The intertonic dataset was taken over from the datasets that I had built in my dissertation (which were taken from Wells (2008)) and then annotated for the variables of interest for this study. We compared the two dictionaries on a random sample of 300 words (100 for each subset) and found that the two dictionaries largely agree with one another and mainly differ in their use of the symbols /u:/ vs. /u/ or in some rare and foreign words. However, as a precautionary measure, the two datasets were always analysed in parallel and were never merged.

Following Hammond (2003), we coded vowel reduction as a four-point scale based on the dictionary data (Full, Full ~ Reduced, Reduced ~ Full, Reduced). If only a full pronunciation is given, the vowel was coded as 1, if only a reduced pronunciation is given, the vowel was coded as 4, and if both a full and a reduced pronunciation are given, vowels were coded as 2 or 3 depending on the order in which these variants are given. For example, the entry in Wells (2008) for the verb *extract* reads as “**ɪk** 'strækt ek-, ək” and so it was coded as 3 (as in the Arab Rule study, /ɪ/ is assumed to represent a reduced vowel in non-primary stressed <e>). We assumed that the symbols /i/ and /u/ represent reduced/unstressed variants of /ij/ and /uɰw/, respectively. In that study, we preserved the symbols /ə/ and /ʌ/, the former being analysed as a reduced vowel and the latter a full vowel, but in the system used here, they are seen as representing the reduced/unstressed and full/stressed realizations of /ə/. We coded the different variables of interest for all the data (nature of the coda, closedness of the syllable, weight of the

first syllable for the intertonic position, spelling, token frequencies, morphology, etc.). Let us just say a word about three of them for the clarity of the presentation of the results below. For SYLLABLESTRUCTURE, we used a three-way distinction, with OPEN and CLOSED but also sC for cases in which the vowel is followed by an /sC/ cluster because such clusters are a notorious issue for syllabification (see e.g. Scheer & Ségéral 2020; Goad 2012). Note that vowels followed by <rC> clusters were coded as CLOSED as they were found to pattern with other closed syllables. In order to code vowel features in non-derived words, we used orthography as it can be taken to indirectly represent distinctions between different classes of vowels. This was done only for monographs as there were too few items for each different digraph, and <i> was excluded as it is not possible to interpret /ɪ/ as either full or reduced (except if it is spelled <e>). Note that this variable, which we called GRAPHEME, was tested in a separate model. Finally, we included a variable FOREIGN that is meant to encode whether or not the word is foreign. We included this variable because, in the first presentations of this study, it was suggested to us that the effects of frequency that we had found (a smaller proportion of reduced vowels in low-frequency words) might be attributed to foreignness: it is possible that, as foreign words are often not very frequent, what appears to be an effect of frequency may be – at least partly – an effect of foreignness. Such an effect might be seen as a phonological difference between the different sublexica that will be described in Ch. 5. As it cannot be assumed that language users know the etymology of loanwords, we must rely on formal characteristics that are available to them, and not etymology. For this, we mainly relied on the characteristics detailed in Dabouis & P. Fournier (2022): word endings appearing almost exclusively in loanwords (e.g. *-i*, *-o*, *-eur*, *-aise*), foreign spelling-to-sound correspondences (e.g. <i> - /ɪj/ in *elite* or *pastis*) and semantics referring to foreign cultures, such as foreign currencies (e.g. *koruny*, *pistole*, *rupee*), food and drinks (e.g. *champagne*, *kebab*, *trepang*), functions (e.g. *hussar*, *savoy*, *ukase*) and objects (e.g. *palankeen*, *pirogue*, *sitar*). This variable was only coded for non-derived words, and about half of the data were identified as foreign in that way: 600/1,235 (49%) in the initial pretonic dataset and 275/474 (58%) in the intertonic dataset.

The data were analysed using R (R Core Team 2023, v.4.1.3) using ordinal logistic regression with VOWELREDUCTION as the dependent variable. This type of regression model is necessary because the dependent variable is a scale. This was done using the `polr` function from the MASS package (Ripley *et al.* 2019). Models were progressively simplified using a step-by-step fashion following standard procedures (e.g. Baayen 2008). We kept only predictors which significantly improved the model, which was measured in three ways:

- the t-value for the predictor variable has to be between -2 and 2;
- the AIC³⁵ of the model including the predictor variable has to be at least two points lower than a model without it;
- a likelihood ratio test comparing the model including the predictor variable and a model without it should have a p-value that is lower than 0.05.

Moreover, the residuals of the final models were analysed using the `resids` function of the `sure` package (Greenwell *et al.* 2018), which uses surrogate residuals (Liu & Zhang 2018), as

³⁵ This stands for the Akaike Information Criterion, which is a measure of the relative quality of a statistical model.

ordinal logistic regression models cannot be analysed directly using common tools for residual analysis.

2.2.2.3. *Results*

The models keeping only the best predictors show effects of SYLLABLESTRUCTURE, SPELLING, LOGFREQUENCY, FOREIGN and, for the intertonic position, WEIGHTS1, which are all consistent with the claims found in the literature, so that full vowels are more common if:

- the vowel is a digraph (e.g. *audition* /*o:*'diʃən/, *araucaria* /*a.ɪo:*'ke:ɪnjə/) than if it is a monograph (e.g. *falafel* /fə'la:fəl/, *pioneer* /*,pajə*'ni:);
- the vowel is in a closed syllable (e.g. *campaign* /kam'peɪn/, *chimpanzee* /*,tʃɪmpən*'zi:/) than if it is in an open syllable (e.g. *gorilla* /gə'ri:lə/, *propaganda* /*,pɒpə*'gandə/). Vowels followed by /sC/ clusters fall in between the two (e.g. *cascade* /kas'keɪd/, *balustrade* /*,balə*'streɪd/);
- the word was categorized as foreign (e.g. *ménage*, *chandelier*) than if it was not (e.g. *molest*, *incantation*);
- the word has a lower frequency than if it has a higher frequency;³⁶
- in the intertonic position, if the first syllable is heavy (e.g. *ecraseur* /*ɛjkɪa:*'zə:/, *trampoline* /*,trampə*'lijn/) than if it is light (e.g. *magazine* /*,magə*'zi:n/, *memorandum* /*,memə*'ɪandəm/).

To illustrate one of those effects, that of syllable structure, the distribution of the data for monographs only (to avoid biases from spelling) is shown in Figure 3.

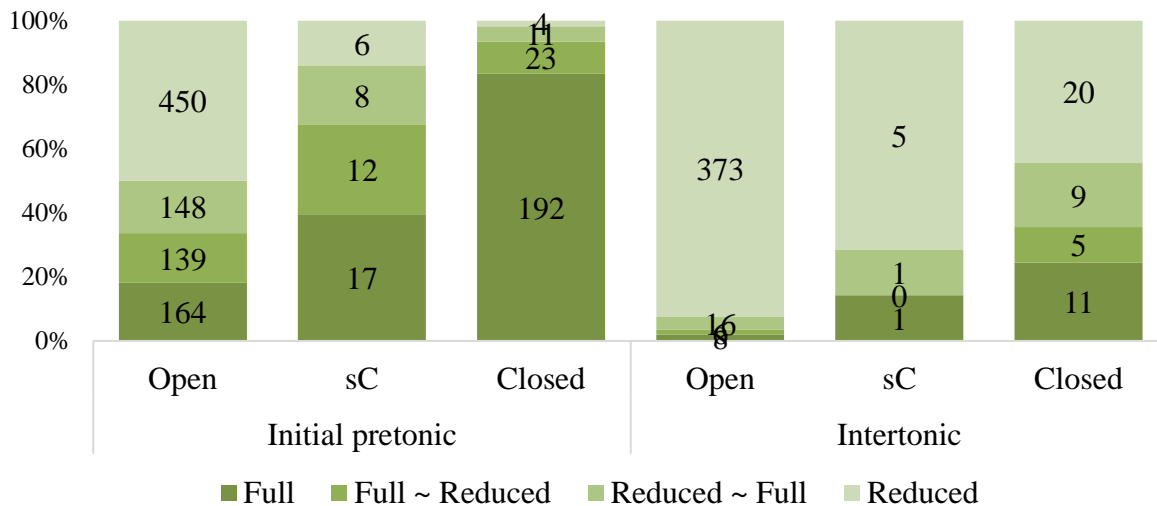


Figure 3. Vowels found for monographs in non-derived words depending on syllable structure

³⁶ There is no statistically significant difference between the frequencies of the words coded as foreign and that of those coded as non-foreign (Mann-Whitney U test; initial pretonic: $W = 164217$, $p = 0.1885$; intertonic: $W = 25911$, $p = 0.70$), and so our results suggest that the effects of foreignness and of frequency are independent from one another.

An additional result that can be seen in Figure 3 and can be observed in all subsets of the data is that full vowels are a lot more common in the initial pretonic position than they are in the intertonic position. Once again, this is consistent with the claims made in the literature.

We fitted another model including the variable GRAPHEME, encoding the difference between the four monographs <a, e, o, u>. We tested it in a separate model because it could be controversial to use orthography as an indirect way to study the potential effects of vowel features on reduction. We did find a significant effect, so that <a> and <e> pattern together and <o> and <u> do too, so that the former two tend to represent reduced vowels more often than the latter two. This can be seen in Figure 4.

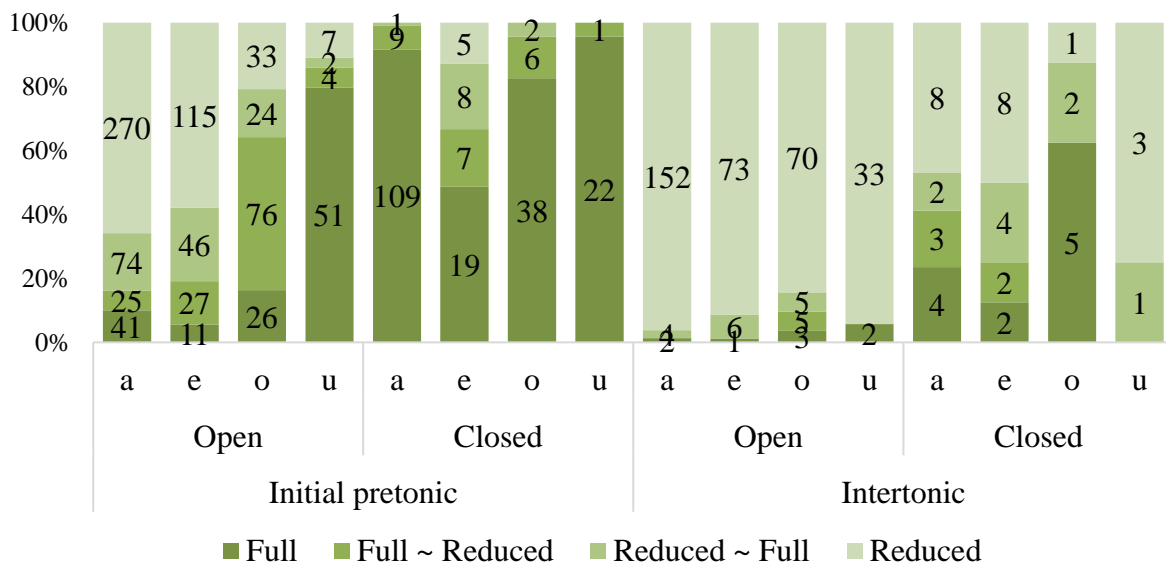


Figure 4. Vowels found for non-derived words depending on syllable structure and grapheme, among <a, e, o, u>

Note that, in closed syllables, <a> patterns with <o> and <u> when it is followed by <rC> and is near-systematically realized as /ɑ:/. Those observations could suggest that back vowels reduce less often than front vowels.

Thus, our results are consistent with several claims found in the literature, notably those on the role of the position of the vowel, of syllable structure and orthography. We found vowels followed by /sC/ to have a behaviour that is between that of open syllables and that of closed syllables, which may suggest that the syllabification of those clusters is variable. They also support the claims regarding the effects of frequency, and we found those effects to be independent from those of foreignness. The evidence for the effect of orthography is relatively weak as we cannot test whether this is a direct effect of orthography or an effect due to the fact that the vowels that are represented by digraphs are different from those that are represented by monographs. We have some evidence in favour of an effect resembling the Arab Rule, as intertonic vowels that are preceded by a heavy syllable are more often full than those that are preceded by a light syllable. As was seen previously, the Arab Rule also involves an interaction with the nature of the coda, but we found no evidence of this in our data, both for place and manner of the coda. We also find a different behaviour for the four monographs investigated which is consistent with the literature too, as Tokar (2019) found that <o> reduces less than

<a> in open syllables in the initial pretonic position. Finally, the effect of foreignness is a new finding that has not been reported elsewhere to our knowledge.

We will revisit some of those effects further on when the results found for the other categories of words will be reported. The results reported here will constitute a useful baseline for the evaluation of the effects of morphology.

2.2.3. *Vowel reduction in French loanwords*

As mentioned previously, in the first presentations of our work on vowel reduction, it was suggested to us that at least part of the effect of frequency on vowel reduction could be attributed to the fact that low-frequency words are loanwords. The underlying assumption was that words that are perceived as foreign would have reduced vowels less often than words that are not perceived as such. This hypothesis was stated explicitly in Dabouis & P. Fournier (2022) and tested and confirmed in the study reported in the previous section.

I have also worked on this issue in another paper with Pierre Fournier which focuses on primary stress placement in French loanwords in British and American English (Dabouis & P. Fournier 2024). The study uses dictionary data from Wells (2008) and includes only entries in which French phonemic transcriptions are given. We found that 50% of our 1,043 entries had final stress in British English and that this figure goes up to 78% in American English. Those results and their implications will be discussed more extensively in §5.3.1 but let us focus here in on one aspect: the full or reduced nature of the vowels found in the last syllable of words in which primary stress is not final. This aspect was only considered for British English, because there is existing work providing figures to which our results can be compared only for that variety. Dahak (2011) studies vowel reduction in post-tonic syllables and provides the rates of words which may contain a full vowel in different positions. One notable difference with the methodology that we adopted is that she treats /ɪ/ as a full vowel while we treated it as ambiguous and excluded it from our counts. Thus, the proportion of full vowels that we report might have been higher if we had adopted her methodology. The comparison between her proportions of full vowels and ours is shown in Figure 5.

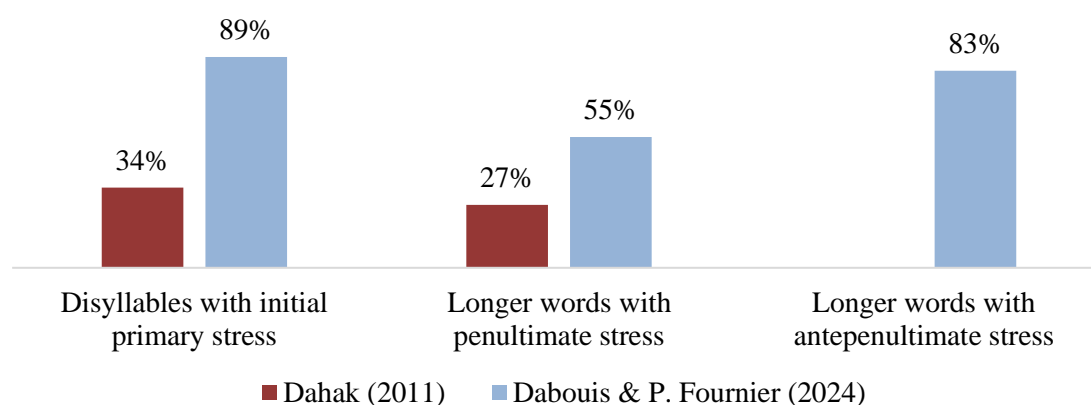


Figure 5. Proportions of words with non-final primary stress which may have a full vowel in their last syllable

As can be seen in Figure 5, in the two positions for which Dahak provides figures, we have considerably greater proportions of words with a final non-reduced vowel among French loanwords than in the whole lexicon. Those results are consistent with the observation made in the previous section concerning the greater tendency for loanwords to have non-reduced vowels as compared to the core vocabulary. However, the fact that we are here dealing with French loanwords suggests that there may be a specificity for that position in that category of words. Indeed, French loanwords tend to have final stress more often than other words, quite likely as a reproduction of the rhythmic group-final stress found in French, and so it is possible that, even though they do not receive primary stress, final vowels in French loanwords are more likely to preserve full vowels for the same reason.

Therefore, possible perspectives for this study include studying loanwords with different source languages to establish whether French loanwords are indeed special, but also using spoken data (from existing corpora or from production tasks) to get a finer sense of how English speakers actually produce those words, and how their knowledge of source languages and possibly their representations of those languages impact their productions.

2.3. Pretonic secondary stress in simplex words

The study of pretonic secondary stress has been the focus of my early career as a researcher as it was the topic of my Ph.D. dissertation (Dabouis 2016b) and I have kept working on this topic ever since. In this dissertation, I dedicated a section to the nature of stress and tried to explain why there are such disagreements in the literature. I had ended up defining the object of my study as syllables which may receive pitch accents, and which are not the most prominent syllable of the word. Now, my views have evolved slightly to those exposed in §2.1.2, and I would now define what I studied at the time as pretonic secondary stress, although now extending this category beyond the classification made by dictionaries increases the number of syllables that would be analysed as having pretonic secondary stress. As mentioned above, the studies on vowel reduction could be rephrased as studies of secondary stress. Therefore, I will revisit some of my earlier results so as to adopt this new view of what stress is.

The general aims that I had for this study was to test the claims that were present in the literature regarding the presence and position of pretonic secondary stresses using a large dataset of dictionary data, as there were hardly any large-scale studies of the issue (one notable exception being Wenszky (2004), with a dataset of 737 words for pretonic secondary stress). Like primary stress, pretonic secondary stress is strongly influenced by morphological structure. I will here only discuss simplex words (and words that pattern with them), and morphological influences on pretonic secondary stress will be discussed in §3.6 and §3.7.2.

In my dissertation and in an overview paper (Dabouis 2020b), I review the literature on pretonic secondary stress. This literature overall agrees that, in words that are not clearly morphologically related to another freestanding word (which I will call ‘non-derived words’), the position of secondary stress largely depends on the following factors:

- A phonological word may not start with two unstressed syllables;
- Adjacent stressed syllables are dispreferred;
- Stress lapses (sequences of unstressed syllables) are dispreferred;

- Heavy syllables attract stress, especially in the second syllable;
- The absence of an initial onset favours second-syllable stress.

I will not here get into the details of those factors. Let it simply be said that they are not all assumed to have the same importance and that the way that they have been formalized in different frameworks varies greatly. Because of the dispreference for adjacent stressed syllables, the number of pretonic syllables (i.e. preceding primary stress) is of great importance in determining possible stress patterns.

The dataset that I used in my dissertation is made up of 5,829 words bearing a secondary stress mark in Wells (2008). All were analysed morphologically, and different morphological classes were treated separately. I had also tested the hypothesis that foreign or learned words may behave differently (see Chapter 5) but found no evidence supporting it. Thus, I had eventually merged all non-derived words into a single dataset.

The dataset contains 146 words which have a single pretonic syllable with a secondary stress mark. Thus, this structure could appear to be relatively rare, but two caveats should be mentioned. First, as the data does not contain words which do not have such a stress mark in the source dictionary, it is impossible to establish the proportion of words with secondary stress on the initial pretonic syllable. Second, the dictionary may show unstressed full vowels in that position, that may be viewed as bearing secondary stress. Thus, the figure that I have just reported may be lower than what should actually be counted. In my thesis, I have suggested that this secondary stress may be a sign of ongoing stress retraction, a claim that is supported by the fact that 48/128 disyllabic words in this subset can also be stressed /10/.

For words with two pretonic syllables, the dictionary transcriptions systematically give the pattern /201(-)/ for all 699 words. If we revisit those results with the assumption that all vowels that are not transcribed as /ə/, /i/ or /u/ bear secondary stress, then we find that a small proportion of words may be analysed as being stressed /221(-)/. As can be seen in Figure 6, those are mainly found in foreign words. If we treat orthographic <i> realized as /i/ as reduced and count the proportion of words which may have a full vowel in their second syllable, we get 11% for §CORE words, 14% for §FOREIGN words and 4% for §LEARNED words (see Chapter 5 on the notation “§”).

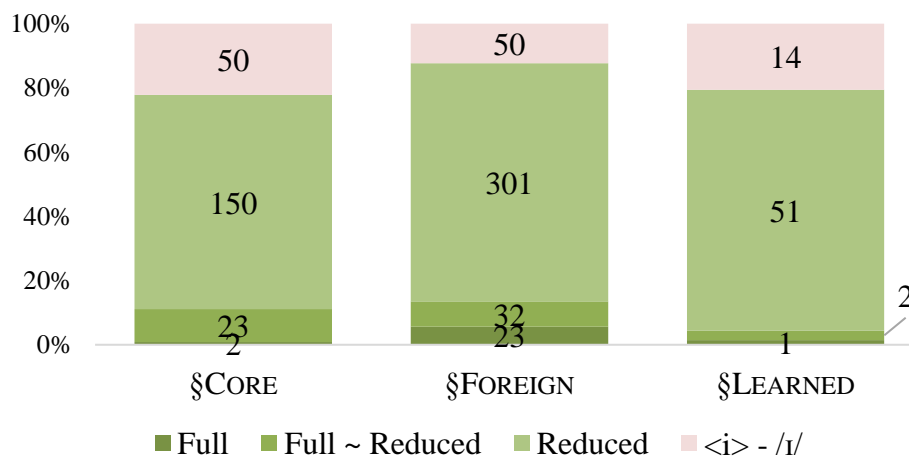


Figure 6. Type of vowel found in the second syllable of non-derived words with two pretonic syllables

Thus, the restriction against adjacent stresses is still relatively clearly enforced, even when we treat all “full” vowels as stressed. /201(-)/ is clearly the regular pattern, an observation further supported by words showing stress variation. There are 36 words in the dataset which oscillate between /201(-)/ and /01(-)/, and the proportion of words with full vowels in the second syllable when primary stress is on the third syllable is only slightly higher (7/36 – 19%) than words without such variation, and those are mostly §FOREIGN words. Such examples are shown in (21).

(21)	importune	/,ɪmpə'tʃʊwn/ ~ /,ɪmpo:-/	~	/ɪm'pɔ:tʃʊwn/
	intercalary	/ɪn'tə:kələɪj/	~	/,ɪntə'kæləɪj/
	composite	/,kɒmpə'zajt/	~	/kəm'pɔzɪt/
	incognito	/,ɪnkɒg'nɪjtəw/	~	/ɪn'kɒgnɪtəw/

As will be seen in §3.6, /201(-)/ is also the favoured pattern in stress-shifted derivatives.

For words with longer pretonic sequences, we expanded the dataset used in my dissertation in Dabouis *et al.* (2017) so as to include proper names, and those results were published in Dabouis (2020b). In the dictionary transcriptions, only six words have two secondary stress marks in the pretonic sequence, but that figure goes up to 29 if all full vowels are treated as stressed. In eight cases (see examples in (22a)), the first two syllables systematically have full vowels, in 12 cases (cf. (22b)) one of the first two vowels may be full or reduced, and in three cases the first and third vowels are (or may be) full (cf. (22c)).³⁷

(22)	a.	Constantinople	/,kɒn,stantɪ'nəwpəl/
		Escorial	/ɛ,sko:ɪj'æl/
		Ticonderoga	/,tɪjkɒndə'ɪəwɡə/ ~ /tɪj,kɒndə'ɪəwɡə/
	b.	Epaminondas	/ɛ,pamɪ'nɒndəs/ ~ /ɪ,pamɪ'nɒndəs/
		Navratilova	/nav,ɪatɪ'ləwvə/ ~ /nəv,ɪatɪ'ləwvə/ ~ /,navɪətɪ'ləwvə/
		Penthesilea	/,pɛnθɛsɪ'lɛjə/ ~ /,pɛnθəsɪ'lɛjə/
	c.	Azerbaijan	/,azəbaj'dʒɑ:n/
		Halicarnassus	/,halɪkɑ:'nasəs/
		Kilimanjaro	/,kɪlɪmən'dʒɑ:ɪəw/ ~ /,kɪlɪmən'dʒɑ:ɪəw/

Based on the discussion in §2.1 and after having listened to several occurrences of those words in speech (using YouGlish), I would now assume that the words in (22a, b) have (or may have) two secondary stresses on their first two syllables, but that there is variability in which syllables receive accents (the first, the second or both). For example, in spoken occurrences of *Constantinople* that we can find on YouGlish, the first two syllables seem to have “level stress” (= two accents?). In *Navratilova* the position of accents and the realization of the first two vowels is difficult to establish auditorily, and for *Halicarnassus*, we never get an accent on the

³⁷ The stress marks indicated in (22) are those given by Wells (2008).

third vowel, which may even be entirely unstressed. Future investigation of those issues using spoken data is necessary to clarify the interaction of stress and accent in those configurations.

Returning to the dictionary data, in Dabouis *et al.* (2017) we were testing the hypothesis that secondary stress is usually left-aligned using an expanded version of my dissertation dataset that includes proper names. This would be a sign of the fundamentally demarcative function of stress, inherited from Old English, as argued by J.-M. Fournier (2007). In that study, based on the secondary stress notation of the dictionary, we had found that secondary stress is best predicted by syllable closedness of the first two syllables and the presence or absence of a word-initial onset. If we revisit the data by assuming that all full vowels are stressed, by sorting syllables into light and heavy³⁸, and if we focus on words with three pretonic syllables, the picture is quite complicated, as can be seen in Figure 7.

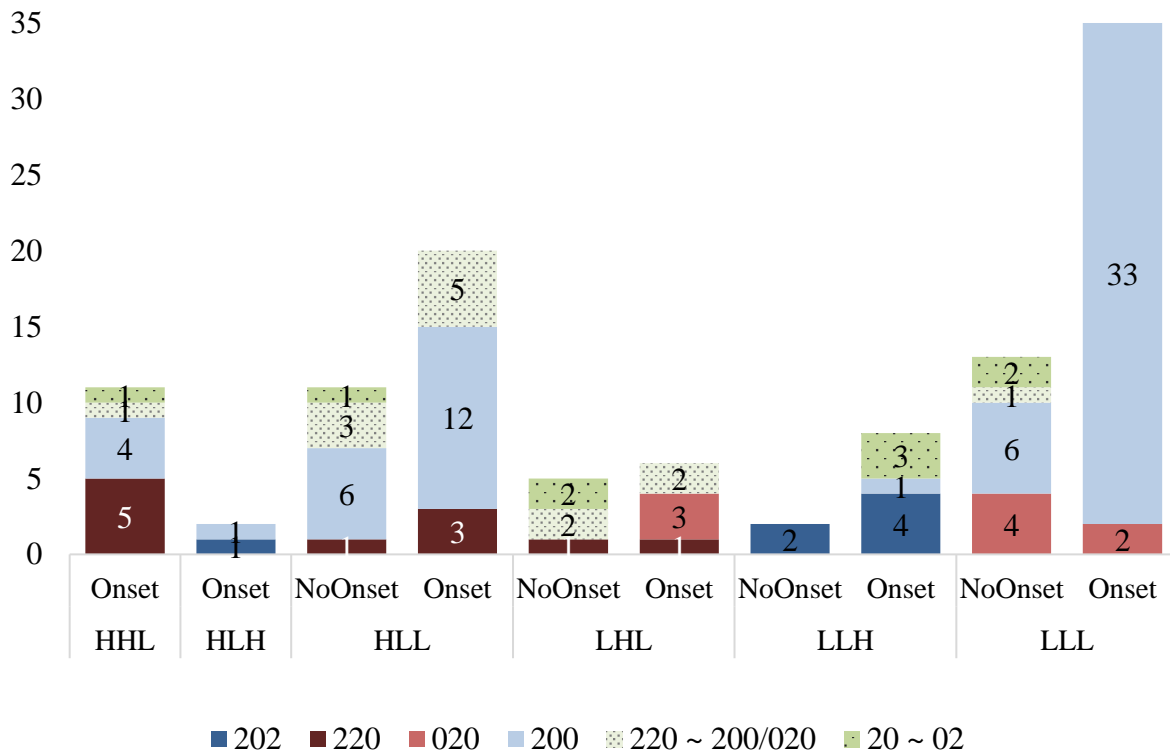


Figure 7. Position of secondary stresses in monomorphemic or deradical words with three pretonic syllables³⁹

The first striking observation is that the clearly dominant pattern is /200-/ (63/113 – 56%) words, which we find in all segmental configurations except HHL and onsetless LHL. The other patterns are never found in more than 12 words, do not appear to be the “rule” in any segmental configuration, and there is overall a great deal of variation. However, second-syllable stress (as /220-/ or /020-/) is possible for all words with a heavy second syllable.

One possible approach to secondary stress placement is to assume that it follows from the left-alignment of moraic trochees. This analysis would predict that LH and HL feet will be

³⁸ Assuming here a rather classical distinction, with light defined as a short vowel in an open syllable and heavy as either a long vowel or a closed syllable.

³⁹ There can be ambiguity on the stress of /əw/ in the dictionary when it bears no stress marks. I am assuming here that it is not stressed if the glide may drop but stressed if it cannot.

avoided because they are trimoraic. Otherwise, syllables will be maximally parsed into minimal feet. If we now compare the predictions made by the moraic trochee approach and the results seen in Figure 7, we get the figures in Table 8.

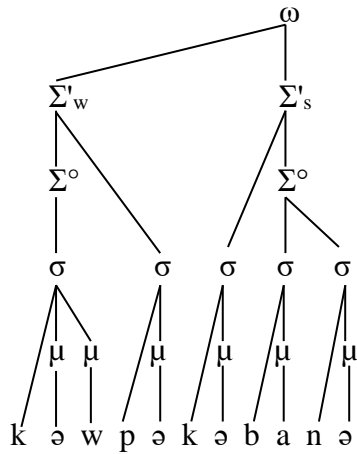
Weight configuration	Expected parsing	Corresponding stress pattern	Predicted pattern observed in data	Variants included
LLL	(LL)L	/200-/	39/48 (81%)	43/48 (90%)
LHL	L(H)L	/020-/	7/11 (63%)	10/11 (90%)
LLH	(LL)(H)	/202-/	8/10	9/10
HLL	(H)(LL)	/220-/	4/31 (12%)	12/31 (34%)
HLH	(H)L(H)	/202-/	1/2	1/2
HHL	(H)(H)L	/220-/	5/11 (45%)	6/11 (54%)

Table 8. Predictions of the position of pretonic secondary stresses according to the moraic trochee approach and the results observed. Brackets represent foot boundaries here.

Thus, 33 words (29%) never have the stress patterns predicted by the moraic trochee approach. Let us have a closer look at those items to see what could explain the discrepancy between predictions and observations.

There are three big types of “exceptions”. First, among LLL words, the exceptions are words that are variably or systematically have second-syllable stress. Seven out of nine of those words have no onset, and thus this absence may be a possible explanation that would be consistent with the available literature. Second, among HHL words, four of the “exceptions” are words which have unstressed occurrences of /əw/ in their second syllable, with optional loss of the offglide (e.g. *Novosibirsk* /,nəwvə(w)sɪ'brɪ:sk/; *Laodicea* /,ləjə(w)dɪ'sɪjə/). It is possible that this loss is very common and thus that they surface as HLL. Finally, HLL is the configuration in which the predictors underperform the most, with 18/31 words which only have the /200-/ pattern. Here the explanation may be theoretical. It is possible that the correct footing in those words is in fact not (H)(LL) but (HL)(L(LL)) as shown in (23) for *Copacabana*, in which the two medial light syllables adjoin to adjacent feet but do not form a minimal foot themselves (see §2.1.2.4).

(23)



If that is the case, then we would expect to see independent evidence for this parsing, i.e. foot-initial behaviour in the third syllable. One word, *mujaheddin* /,mʊwdʒəhɛ'dɪjn/⁴⁰, indeed has an /h/ in the relevant position, which would be consistent with a foot structure such as that in (23). Five words have voiceless stops in that position (*entrepreneur*, *intrapreneur*, *Copacabana*, *Semipalatinsk*, *Yoknapatawpha*) and so acoustic analyses could establish whether or not they are aspirated, as would be predicted by (23). Thus, it is possible that what appear to be exceptions are indications of constraints on footing that go beyond constraints on foot size and alignment.

To conclude, let us mention a few possible reservations on those results. When we created the dataset described in that section, we had tremendous difficulties with delineating where morphologically complex words end and where simplex words begin. We left out words which are probably related to an existing form, although not through any kind of productive morphology (e.g. *comédiénne* ↔ *comédian*, *Illùminàti* ↔ *illùminate*), words with recurring elements that may be perceived as formatives (e.g. *Àbergavénny*, *Àbertilléry*; see §3.3.2) and toponyms in *-(i)stan* and *-abad* were left out as they seem to function like autostressed neutral suffixes (e.g. *Waz̀ristán* ↔ *Waz̀ir*; *Jallàlabád* ↔ *Jalál*). This points to the fact that, unsurprisingly, simplex words with long pretonic strings are difficult to come by, and they are often quite obscure words, proper names, and quite foreign. It is quite likely that they are not used very often, and English natives would probably often be at pains to read those words.⁴¹ In a joint presentation with Guillaume Enguehard and Nicolas Lampitelli (Dabouis *et al.* 2019), we explored the idea that language users tend to analyse such words as having some kind of internal structure (which we called ‘crypto-morphology’). This intuition seems to be shared by others such as James L. Fidelholtz (p.c.):

I suggest that all long words in English that are unanalyzable morphologically (e.g. *cassabanana* – a fruit completely unrelated to either *cassaba* or *banana*) are broken by folk etymology into manageable (‘bite-sized’) chunks separated by a word boundary.

⁴⁰ However, this word has a main pronunciation that would entail a “degenerate” monomoraic foot: /,mʊwdʒəhɛ'dɪjn/.

⁴¹ I have myself witnessed this as I have used some of them as distractors in an exploratory production study of /ɹ/ gemination (see §3.7.3), in which participants did not know most of the words and really struggled with producing them.

We find similar arguments in Tokar (2018) on the stress of *abracadabra*, or Flynn (to appear), who explores the idea that such long words are parsed into two phonological words.⁴² In our joint presentation, we tried to substantiate that claim and tried to capture the effects usually attributed to foot structure in the framework of Government Phonology but, at this stage, the analysis is not convincing enough to be published. There too, valuable information may come from phonetics because, if there is really a (phonological) word boundary between the two ‘crypto-constituents’, then we would expect to see other manifestations of that boundary (e.g. pre-boundary lengthening).

⁴² This is probably also what Prince (1983) hints at when he discusses “felt” feet which, according to him, follow the following generalization: “a weak syllable, not immediately after the stress, will procliticize to a following (main-stressed?) foot”, although “closed syllables seem to resist”. He notes that “the process thus appears to reflect constraints on word form”.

CHAPTER 3

The interaction between morphosyntax and phonology

Most of the processes that I have been working on involve – at least partly – some influence of morphology on phonology. This is the reason why this chapter is the longest of this document. I will start by presenting my approach to morphology (§3.1) and how theories incorporate morphosyntactic information into the phonology (§3.2). Those first two sections will introduce concepts, definitions and theoretical tools which will be useful in subsequent sections, especially to readers who are not familiar with those issues. Then, I will discuss work that I have been conducting on quite obscure types of morphological structures and their phonological manifestations (§3.3), which have not been studied in much detail as most studies of the morphology-phonology interface have focused on productive morphology. English is well-known for having a stress system that is sensitive to syntactic categories, and so some of my work has focused on verbs and on stress in disyllabic verb-noun pairs, as will be seen in §3.4. The next four sections deal with the phonological behaviour of affixed words. I will start by discussing the issue of suffix classes (§3.5), which is the first question that I ever worked on. Then, I will mention some of the main results from my dissertation concerning the location of pretonic secondary stress in stress-shifted derivatives (§3.6). In the past twenty years, there have been a number of studies testing what is sometimes called the “segmentability hypothesis”. Thus, I will introduce that hypothesis and discuss some of the studies that I participated in which have tested it (§3.7). Finally, I will deal with the issues of locality and containment in the debate surrounding the choice of phonological bases for morphologically complex words and what my contributions on those issues are (§3.8).

3.1. Approach to morphology

The framework that I was trained in is a ‘piece-based’ model, like most phonological models. Guierre (1979) makes distinctions between “morphemes” and what he called “pseudo-morphemes” (e.g. *conduct*, *detain*, *project*) and “quasi-morphemes” (e.g. *anthropo-*, *psycho-*, *-log-*), but no clear definitions of those categories are provided. Among the category of morphemes, Guierre includes affixes that many would not treat as any kind of morphologically-relevant unit such as *obtain* or *select*. I quickly became aware that such analyses were controversial and that, if they were to be upheld at all, that position would have to be backed up by solid evidence. Over the years, I have read extensively about the different approaches to morphology and how different models have dealt with the issues that the traditional notion of the morpheme as a minimal sign faces. In those readings, I found out that the very definition of

morphology is subject to disagreements, as we can find a series of different definitions listed in (24), and taken from Stewart (2016: 1).

- (24) a. The set of “constructions in which bound forms appear among the constituents” (Bloomfield 1933: 207)
- b. “The study of the word formation processes of language” (Siegel 1979: 12)
- c. “The syntax of words” (Selkirk 1982)
- d. “The component of grammar that builds words out of pieces” (Pinker 1999: 293)
- e. “The complex process by which abstract morphosyntactic representations are realised morphophonologically” (Aronoff 1994: 9)
- f. “The study of morphemes and their arrangements in forming words” (Nida 1949: 1)

Therefore, it seems necessary to begin this section with my approach to morphology, its nature and basic units, and to introduce the terminology that is required to understand the following sections. I will start by discussing the basic units of morphology (§3.1.1) and will then turn to the issue of morphological relatedness (§3.1.2).

3.1.1. *Basic units*

The relationship between form and meaning in language is deemed to be largely arbitrary, but that is not the case in morphologically complex words (Stevens & Plaut 2022) as recurrent associations of form and meaning can be found in different words (e.g. *bedroom*, *bedding*). There is a great deal of diversity in linguistics as to how such associations are described, and a central part has to do with the unit taken to be the central unit of morphology: the word (or lexeme) or the morpheme (or morph)?

Before discussing different theoretical approaches to morphology, it seems necessary to start with some basic definitions. While the notion of ‘word’ is particularly problematic and has mostly eluded a universal definition (although see Haspelmath (2023) for a recent proposal), morphologists usually use the terms word-form and lexeme to describe the units that are informally referred to as ‘words’ (see e.g. Aronoff & Sims 2023; Bauer *et al.* 2013: 8-9; Haspelmath 2023; Stewart 2016: 2). The **lexeme** is often described as a ‘dictionary headword’ or, more formally, as “an abstraction over one or more word-forms that convey the same lexical meaning”, while a **word-form** “refers to a phonological/orthographic shape irrespective of meaning” (Bauer *et al.* 2013: 8). For example, the examples in (25a) are lexemes, and their corresponding word-forms are shown in (25b).⁴³

⁴³ It is often noted that this distinction relies on another, that between derivation and inflection, which is also a controversial issue in morphology as no precise boundary can be identified between the two. To stay focused on the main issues that are relevant to the work that I have conducted, I will not get into this additional issue, but the reader may consult Anderson (1992: §4.1), Bauer *et al.* (2013: Ch. 24) or Haspelmath (2002: Ch. 4) for discussions of the topic.

(25)	Lexemes	Word-forms
	a. BE	b. am, is, are, was, were, being, been
	LOVE	love, loves, loved, loving
	CAT	cat, cats

When it is not crucial to distinguish lexemes from word-forms, I will keep using the term word.

Units that are found below the word level are the source of even stronger controversies. In textbooks, we often find the notion of **morpheme** which is classically assumed to be a minimal meaningful unit or minimal sign, some kind of indivisible atom that is involved in the construction of words (Bauer *et al.* 2013: 13-14; Haspelmath 2002: 16; Katamba & Stonham 2006: 20). Moreover, a morpheme would be a form of abstract unit that is realized concretely by **morphs**. In cases where a morpheme is realized by different morphs, those are called **allomorphs**. We can further distinguish cases where the form of the allomorph is predictable on phonological grounds (e.g. /{z, s, ɪz}/ for the plural -s) from **suppletion** where it is not (e.g. *good – better – worse*).⁴⁴ However, there are many issues for this traditional approach to the morpheme. Here is a very short summary of the questions that it raises, taken from a very rich literature on the issue:

- If different morphs realize the same meaning or function (e.g. plural in *dogs* vs. *oxen*), should we always assume that they are (suppletive) allomorphs?
- How should we analyse cases in which no clear meaning can be identified despite the form being recurrent (e.g. *deduce*, *induce*, *reduce*, theme vowels) or found in combination with more ‘prototypical’ morphs (e.g. *cranberry*)?
- How should we analyse cases in which a morphosyntactic feature has no phonological/orthographic correspondent? Should we posit ‘zero-morphs’ (e.g. is the singular for *dog* to be analysed as $[[dog]_{\text{root}} - \emptyset]_{\text{N}}$)?
- How should we analyse morphs that are associated with several morphosyntactic features (‘cumulative morphs’, e.g. Lat. $-\bar{o}$ marks both the first person singular and the present tense)?
- Should word-forms always be exhaustively analysed into morphs (this has been called “total accountability” by Hockett (1947))?

To this series of questions whose answers are far from obvious, we can add that there is considerable terminological confusion around the term ‘morpheme’. As pointed out by Haspelmath (2020), it can refer to a minimal form, to “a set of minimal forms with identical syntacticosemantic content” or to “a minimal element of (morpho-)syntactic representation”. He instead argues for the use of the term ‘morph’, defined as minimal linguistic forms (i.e. “a minimal pairing of syntacticosemantic content and a string of phonological segments”), which are segmental (i.e. there are no zero-morphs) and continuous (i.e. there cannot be discontinuous morphs, e.g. circumfixes).

⁴⁴ Haspelmath (2020), who argues for the use of the term ‘morph’ and not that of ‘morpheme’, proposes to call those cases “variomorphy” and “suppleomorphy”, respectively. We can note that the distinction between those two categories is very theory-dependent, as what is considered predictable phonologically may vary from one framework to another.

The controversy surrounding the concept of morphemes has led to a split between two types of approaches to morphology. First, **word-based approaches** (or lexeme-based approaches) assume that word forms (or lexemes) are the basic units. Models differ in their treatment of recurrent word parts, with some assuming that they are abstractions over full forms which are not themselves listed in the lexicon, while others do not give them any role. Early proponents of word-based morphology are Aronoff (1976) and Booij (1977), and some more recent models are A-morphous Morphology (Anderson 1992), Construction Morphology (Booij 2010, 2012; Booij & Audring 2017) or Relational Morphology (Aronoff & Sims 2023). Second, **morpheme-based** approaches (sometimes also called piece-based approaches) assume that morphemes are the basic units of morphology, and that they are the units involved in novel word formation. Some proponents of such models are Selkirk (1982) or Taft (2004), and the most widespread morpheme-based morphological theory today is probably Distributed Morphology (Halle & Marantz 1993). This theory assumes that morphology is a part of syntax, and adopts the Separation Hypothesis, according to which “the semantic and phonological facets of the morpheme must not be represented in the same primitive object” (Embick *et al.* 2022). This means that it is no longer necessary to have one form associated with one meaning: morphemes are assumed to be abstract units carrying a bundle of abstract features (e.g. [\pm past], [\pm plural]) and which are “assigned phonological values by the grammar” (Marantz 2013). They might be realized by a variety of phonological “exponents”, and “each phonological exponent might realise a number of different morphemes” (*ibid.*).

The impression that I get from reading various strands of the literature is that certain research communities have largely adopted one approach over the other: nowadays most morphologists seem to adopt word-based approaches while phonologists and psycholinguists tend to adopt morpheme-based approaches. As a phonologist, it seems difficult for me to take a stand on this divide, as there are many convincing arguments in favour of word-based models, but most phonological theories assume morpheme-based approaches (although this is not always explicit). Thus, I will mostly avoid the term ‘morpheme’, except occasionally when citing previous work which uses it. The main case in which I will need to refer to recurrent word parts will be when I discuss the phonological properties of words which contain such recurrent parts, but that those parts have little or no identifiable meaning. Therefore, following Bauer *et al.* (2013: 16), I will refer to such elements as **formatives**, defined as “an overarching category that includes both morphemes and elements contributing to the construction of words whose semantic unity or function is obscure or dubious”.

The distinction between morphemes (minimal signs) and formatives with no clear meaning is certainly not clear-cut, and I am increasingly convinced by a more gradient approach to morphology. As Plag & Balling (2020: 310) put it,

It seems that a gradient view of morphology in which speakers and listeners make use of all sorts of sound-meaning pairings can better account for the intricate experimental results. These results suggest a continuum of variable association strengths between sound patterns and meaning. At one end of the continuum we find the constellation of strong and clear associations that is traditionally referred to as ‘morpheme’, while towards the other end of the continuum association between meaning and form becomes less and less strong and consistent. [...] Theories that restrict their inventory of units to morphemes must either stretch this notion to such an extent that it becomes unrecognizable, or they simply ignore

important evidence that morphemes are not the neat discrete units that these theories are working with.

Therefore, the traditional, ‘prototypical’ morpheme (one form – one meaning) is one configuration among many of the various possible associations between form and meaning. One way to represent some of the different types of configurations that we can find along the gradients of shared form and meaning across lexemes might be something like Figure 8.⁴⁵ Examples for those configurations are given in (26).⁴⁶

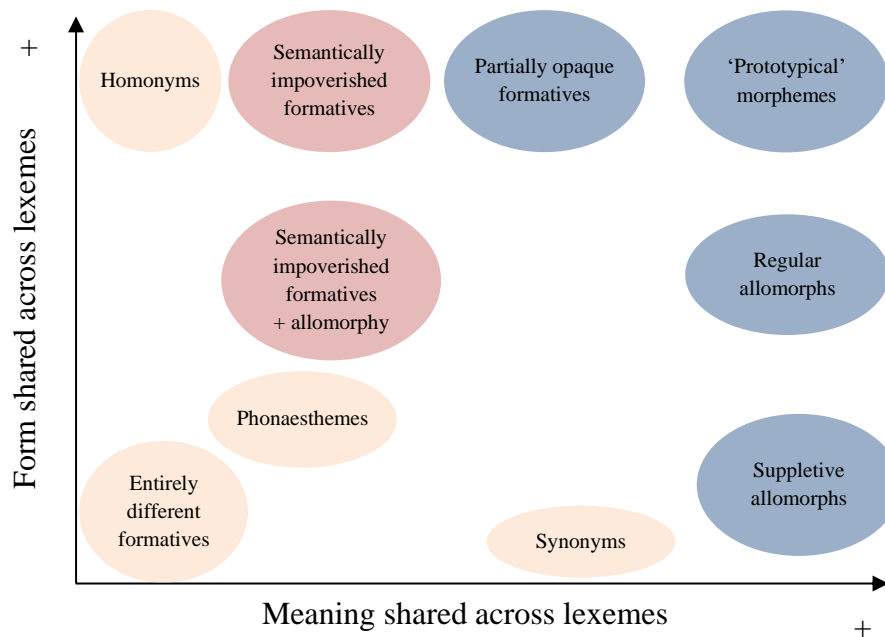


Figure 8. Types of configurations with different associations of form and meaning (units that are typically assumed to be morphemic are shown in dark blue, those that are controversial are shown in red, and those that are usually not assumed to be part of morphology are shown in light orange)

- (26) a. Homonyms *bat, trunk, park, bank, rock, bark...*
 b. Phonaesthemes */gl/ for light in glow, glisten, glimmer...*
 c. Synonyms *bad, awful, terrible, horrible...*

⁴⁵ Note that Figure 8 might be taken to suggest a form of scalar organization rather than a gradient one. The types of units shown here are meant to represent points on a gradient, and the separation between is certainly not clear-cut but indeed gradient.

⁴⁶ One additional dimension that should be considered is distributional recurrence. As will be seen in §3.3, it has been argued to be a key element in the recognizability of semantically impoverished formatives.

d.	Semantically impoverished formatives	Bound roots: <i>ject, port, tract, sist...</i> Obscured affixes: <i>-ish_V, a-, be-...</i> Compound constituents: <i>greenhouse, strawberry...</i> Derivatives that have drifted from their base: <i>universe - university</i>
e.	Semantically impoverished formatives + allomorphy	Bound roots: <i>duce</i> /{djʊs, dəkt, dəkf}/ <i>cede</i> /{sɪd, səd, sɛs, sef}/ Obscured affixes: <i>con-</i> /k{ə, ɔ}{n, m, Ø}/ <i>ob-</i> /{əb, ɒb, ə}/
f.	Partially opaque formatives	<i>roughly</i> : ‘without being exact or fully authenticated, approximately’ <i>government</i> : ‘the governing body of persons in a state, community, etc’
g.	Regular allomorphs	plural /{s, z, ɪz}/, negative <i>in-</i> /{ɪn, ɪm, ɪl, ɪz}/...
h.	Suppletive allomorphs	<i>bad</i> ♦ <i>worse</i> , <i>good</i> ♦ <i>better</i> , <i>be</i> ♦ <i>was</i> ...
i.	‘Prototypical’ morphemes	<i>cat, sweet, baby, member, squeeze, -ness</i> ...
j.	Entirely different formatives	<i>dog – narrow, open – theory, canal – listen</i> ...

Finally, there are three types of basic units that we need to define. First, **roots** correspond to minimal acategorical items which may be bound or free (Bermúdez-Otero 2018b; Giegerich 1999). A root is generally assumed to have lexical content that denotes a thing, an action or a property, and is what remains when all affixes are removed (Bauer *et al.* 2013: 17; Haspelmath 2020).⁴⁷ I will treat formatives such as (26d, e) as bound roots, even if their lexical content is sometimes very impoverished. Second, the term ‘**stem**’ is used in a variety of ways, sometimes synonymously with ‘root’, sometimes with ‘base’, and sometimes to designate the part of a word-form that remains when inflectional affixes are removed. I will only be using this term when citing previous work, and so there will be three main uses of the term:

- In the psycholinguistic literature which will be discussed in §3.3.1.3, where it seems to correspond to what we have just defined as a root;
- In Stratal Phonology, where it is defined as “a lexical item specified for syntactic category (N, V, A, etc.). In certain cases, it is necessary to distinguish further between **derivational stems** and **inflectional stems**, where a derivational stem belongs to a syntactic category but must undergo some further morphosyntactic operation before it becomes inflectable” (Bermúdez-Otero 2018b; emphases in the original);
- Raffelsiefen, whose work will be mentioned on Output-Output Correspondence (§3.2.3) and affix classes (§3.5) uses the term to designate “a specific part of a morphologically

⁴⁷ Even though it is acategorical, Haspelmath notes that the lexical content that a root bears often determines which syntactic category it will tend to adopt (e.g. if it denotes a thing, it will tend to be a noun).

complex word, namely the part that remains when one or more affixes are stripped off” (Raffelsiefen 2023). This includes derivational affixes, and so a stem should be differentiated from a base. For example, the base of the word *decisive* /dɪ'sɪjɪv/ is *decide* /dɪ'sɪd/, but its stem is *decis-* /dɪ'sɪjz/ (example from Raffelsiefen p.c., with my own transcriptions).

Finally, **affixes** are “obligatorily bound items which attach to roots” (Bauer *et al.* 2013: 17). As with roots, some of them are quite productive and could be claimed to be prototypical morphemes (e.g. *co-*, *mis-*, *-ness*, *-ly*) but others, such as those in (26d, e), which most morpheme-based morphologists would probably not treat as morphemes, will be treated as semantically impoverished affixes. The arguments for doing so will be developed in §3.3.

3.1.2. Morphological relatedness

Let us turn to the terminology required to designate different types of groups of words that share a morphological property or relationships between words. I will mainly be focusing on affixed words, which are seen as the result of **derivation**. The unit to which an affix is attached will be called a **base**, and the result of derivation will be called a **derivative**. In some cases, I will refer to previous work in which we distinguished ‘derived’ words, which have a free base, from ‘non-derived’ words, which have no structure or have a bound base. The latter may also be called ‘deradical’ derivatives.

I will be making little use of the term **paradigm**, apart from occasional mention when referring to work which does use the term. This term refers to “relations among word forms at a level of representation that abstracts away from particular exponents” (Aronoff & Sims 2023), and is often used to refer to “the exhaustive set of inflected forms that share a root or stem” (Albright 2011). As this term has also been used for derivational morphology, and is sometimes used to designate the set of words that share a root or stem (here used in the sense of a lexical item specified for a given syntactic category) or the set of words that share an affix, I will instead be using two different terms. Following Bauer *et al.* (2013: 19), among many, I will be using the expressions **morphological category** to refer to “all words derived with the same affix or by the same morphological process” (e.g. all the words with the adjectival suffix *-ic*) and the **morphological family** of a word to refer to “all morphologically complex words (i.e. types) in which that word occurs as a constituent. For example, the morphological family of *accept* would contain *accept*, *acceptable*, *acceptably*, *acceptability*, *acceptance*, *accepter/acceptor*, *accepting*, *acceptive*” (*ibid.*: 67).

In the following sections (especially §3.8), I will be dealing with what Bermúdez-Otero (2016) calls “paradigmatic dependencies” and defines as in (27).

(27) **Paradigmatic dependency in morphophonology**

The form of a linguistic expression *a* is predictable from the surface representation of one or more morphosyntactically related expressions {*b*, *c*, ...}.

As we will see, paradigmatic dependencies within morphological families have often been expressed with various theoretical tools involving faithfulness between a word and another

member of its morphological family. Thus, as pointed out by Steriade (1997), it is important to distinguish between the different ‘bases’ which a word may have, notably the morphosyntactic base and the phonological base. In certain cases, it has been argued that paradigmatic dependencies may hold between a derivative and another member of its morphological family which is not its morphosyntactic base. For all cases that will be discussed, the morphosyntactic base is an immediate subconstituent of the derivative, and so I will call it the **local base** (following e.g. Stanton & Steriade 2014). To refer to any other member of the morphological family which may act as a phonological base, the term **remote base** is often used in the literature on lexical conservatism (§3.5.6). I will use this term mainly when discussing lexical conservatism, and will otherwise prefer the term **non-local base**. As part of the controversy between different theories has to do with the distinction between bases that are subconstituents of the derivative and bases which are not, I will distinguish between **distant bases**, which are lexemes that are ‘contained’ within the derivative (e.g. *direct* ♦ *director* ♦ *directorial*), and **co-derivatives**⁴⁸, which are lexemes which are members of the morphological family of the derivative and which are not contained within it (e.g. the local base of *acceptation* is *accept* and its co-derivatives are *acceptable*, *acceptability*, *acceptance* and *acceptor*). Examples for all the terms that we have just seen are shown in Figure 9.⁴⁹

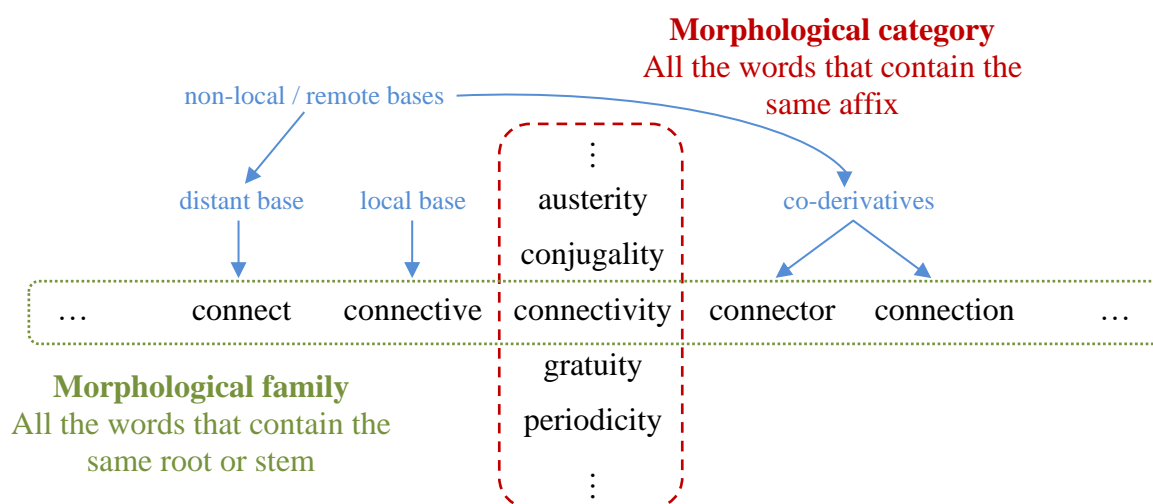


Figure 9. Illustrations of the terminology used to refer to morphologically related words

⁴⁸ The term may also be used to refer to words which share a suffix (Stewart 2016: 2), but here it will only be used to refer to words within the morphological family.

⁴⁹ Let us say a word about the theoretical devices that are used to capture phonological uniformity across sets of morphologically related words, and which will be detailed in the following sections. Within morphological families, phonological uniformity is usually attributed to dependencies between lexemes (achieved through devices such as cyclicity or Output-Output Correspondence constraints) while within morphological categories referring to affixed words, phonological uniformity is typically attributed to idiosyncratic properties of affixes (e.g. stratal affiliation, affix class, cohesive behaviour, segmental makeup). Such distinctions are very dependent on the theoretical framework that one adopts, and one may see them as arising from differences in the structural characteristics of morphological families and categories: there is more form and meaning shared between the members of the same morphological family than between those of the same morphological category, as in the latter only the affix is concerned. Thus, in frameworks in which the distinction between bases and affixes are not relevant, such as analogy-based frameworks (e.g. Skousen 1989, 2009), differences in phonological uniformity within morphological categories and morphological families may be seen as arising from those different structural characteristics.

Finally, following Bauer *et al.* (2013), I am using the symbol ♦ to signal morphological relatedness without necessarily indicating a directionality even if, as will be seen in the coming sections, the privileged directionality for paradigmatic dependency is from the (local) base to the derivative.

3.2. How phonological theories incorporate morphosyntactic information

In generative theories of the morphosyntax-phonology interface, two channels have been used to incorporate morphosyntactic information into the phonology: the procedural channel and the representational channel (Bermúdez-Otero 2011; Scheer 2011: §5). This is what Scheer (2011) calls “Interface Dualism”. **Procedural** communication (also called “derivational” communication in Scheer & Cyran (2018a)) consists in the morphosyntactic control over which parts of the structure are visible to phonological computation at a given stage. In cyclic models, the computation proceeds from the smallest constituent of an expression and outwards onto the full expression. For example, the computation of the phonological representation for *originality* goes through a first cycle for *origin*, a second one for *original* and a third one for *originality*. In Output-Output-correspondence (OO-correspondence) models, the morphosyntax allows for the phonology to see the surface form of morphological relatives during the computation of the phonological representation of a word. **Representational** communication consists in the insertion of phonological objects (e.g. prosodic units, boundaries) into phonological representations which are positioned in reference to the edges of morphosyntactic units. Both channels may be used to account for morphosyntactically induced **misapplication** of phonological processes, i.e. the different phonological behaviour of complex expressions. Misapplication means that a process apparently does not apply in its usual environment (it is said to be **opaque**): it can either **underapply** if it fails to apply in its usual environment or **overapply** if it applies outside of its usual environment. The choice of one channel or another is not always obvious to make. Bermúdez-Otero (2011) even asserts that

The uncertainty whether a particular instance of morphosyntactic conditioning in phonology should be analysed procedurally or representationally is in fact one of the most serious and recurrent obstacles faced by empirical research into the morphosyntax-phonology interface.

The representational and procedural tools that have been used in different theories often differ, although some tools are more used than others, as shown in Table 9.

Theory	Representational effects	Procedural effects	Sample reference
<i>SPE</i>	boundary symbols (+, #)	the cycle	Chomsky & Halle (1968)
Lexical Phonology	prosodic units (built by rules)	the cycle (with levels)	Booij & Rubach (1984)
Stratal Phonology	prosodic units (controlled by ALIGN)	the cycle (with levels)	Bermúdez-Otero (2018b)
Classical OT	prosodic units (controlled by ALIGN)	OO-correspondence	Raffelsiefen (2005)
Strict CV Phonology	empty CV units	the cycle (phases)	Scheer (2012b)

Table 9. The two types of morphosyntactic conditioning acknowledged throughout the history of generative phonology (adapted from Bermúdez-Otero 2011)

This section will briefly describe the last three theories in Table 9, which are the most commonly used nowadays, as those theories will be mentioned in subsequent sections. As Prosodic Phonology is a theory used to capture morphosyntactic effects representationally in both Stratal Phonology and Classical OT, it will be given its own section.

3.2.1. Prosodic Phonology

Prosodic Phonology is “certainly the most influential interface theory of the past 40 years” (Scheer 2011: 303). It has its roots in Metrical Phonology (see e.g. Hayes 1981; Liberman & Prince 1977) and is now commonly used by many phonologists. This theory is fully exposed in Nespor & Vogel (1986), although some its main concepts had been laid down in earlier work (e.g. Booij 1983; Selkirk 1980b). This theory posits that there is a set of prosodic constituents (or categories) that make up the **Prosodic Hierarchy** shown in (28). The exact constitution of the Prosodic Hierarchy varies somehow as some authors do not include the mora or add additional constituents such as the colon or the clitic group.

(28) The Prosodic Hierarchy

Mora (μ)
 Syllable (σ)
 Foot (Σ)
 Phonological (or prosodic) word (ω)
 Phonological phrase (φ)
 Intonational phrase (ι)
 Phonological utterance (υ)

According to Nespor & Vogel (1986), those constituents have four types of motivation:

- there are phonological processes that refer to them in their formulation;
- there are phonological processes that have them as their domain of application;
- they are domains of phonotactic restrictions;
- they bear prominence relations.

We have already seen in §2.1.2.4 how this could be applied to the foot. Thus, this theory is a theory of phonological domains, and the rules that refer to those domains can be organized into three categories, which were initially defined by Selkirk (1980a):

- **Domain span rules**, which apply only if their environment is contained within a given prosodic domain:

$$A \rightarrow B / {}_{Di}(\dots X_Y \dots)_{Di}$$

- **Domain limit rules**, which refer to the edge of a prosodic constituent:

$$A \rightarrow B / {}_{Di}(X_Y \dots)_{Di}$$

$$A \rightarrow B / {}_{Di}(\dots X_Y)_{Di}$$

- **Domain juncture rules**, which apply to segmental strings that are adjacent to a boundary separating two prosodic constituents of the same type:

$$A \rightarrow B / {}_{Di}(\dots {}_{Dj}(\dots X_Y)_{Dj} {}_{Dj}(\dots Z \dots)_{Dj} \dots)_{Di}$$

$$A \rightarrow B / {}_{Di}(\dots {}_{Dj}(\dots X)_{Dj} {}_{Dj}(\dots Y_Z \dots)_{Dj} \dots)_{Di}$$

With regards to the morphosyntax-phonology interface, the key aspect of the theory is the Indirect Reference Hypothesis, which posits that prosodic structure is related to, but **non-isomorphic** with morphosyntactic structure. There is thus a “translation” process during which the surface morphosyntactic structure is mapped onto prosodic representations. This is in line with modular theories of grammar, in which each module is assumed to have its own vocabulary, and so the phonology should not be able to see morphosyntactic categories. However, the Prosodic Hierarchy is quite heterogenous in how it emerges.⁵⁰ Nespors & Vogel (1986: 107) note that only the phonological word and higher categories are “constructed on the basis of mapping rules that make substantial use of morphological notions”, i.e. in a ‘top-down’ fashion. In contrast, moras, syllables and feet are seen as projections based on the properties of segments and are thus constructed in a ‘bottom-up’ fashion. Note that Raffelsiefen (2005) amends Nespors & Vogel’s model in proposing that the phonological word is the lowest constituent which is **necessarily** aligned with a morphosyntactic boundary. This leaves open the possibility for alignment between morphosyntactic boundaries and lower categories.

As we have already seen several uses of the Prosodic Hierarchy for feet in previous sections, let me conclude this section with an illustration for the phonological word. Bermúdez-Otero (to appear) analyses the condition in which the process that he calls “/aI/-raising” applies. This process raises the PRICE vowel in many varieties of English to what he transcribes as [aI] before voiceless consonants. However, he argues that the process applies only under certain prosodic conditions. It applies if the trigger is “in a weak branch of the lowest prosodic node dominating both trigger and target”, which includes the coda, the onset of a following weak syllable or the onset of a following weaker foot (29a). However, it does not apply if the trigger is in the onset of a following stronger foot (29b) or if it belongs to another phonological word. This can happen in compounds (29c), whose constituents have long been analysed as having their own phonological word (see e.g. Bermúdez-Otero 2011; Booij & Rubach 1984;

⁵⁰ This heterogeneity is a source of criticism from Scheer (2011: §374): “the Prosodic Hierarchy is a blend of heterogeneous phonological objects, rather than a homogeneous arboreal representation”.

Raffelsiefen 2007), and transparent prefixed words (29d), whose prefixes have also been analysed as having their own phonological word (see e.g. Booij & Rubach 1984; Raffelsiefen 1999; Wennerstrom 1993).

- (29) a. cíte ('sɛit)
 cýcle ('sɛi.kəl)
 nítràte (('nɛi)Σ(,tɛɪt)Σ)ω
- b. citation ((,sai)Σ('teɪʃən)Σ)ω
- c. hígh schòol (('haɪ)ω(,sku:l)ω)ω'
 tíe shòp (('tai)ω(,ʃɑp)ω)ω'
- d. bì-centénial ((,baɪ)ω(sɛn'tɛniəl)ω)ω
 trî-syllábic ((,tɛɪ)ω(sɪ'læbɪk)ω)ω

The phonological word is a tool which I have used in several analyses, mainly as a tool that is useful to capture the string of behaviours that is associated with semantically transparent prefixes. In my dissertation and in two subsequent papers in particular (Dabouis 2018, 2020b), I discuss characteristics that have been noted in the literature and which I confirmed using a large dataset: the presence and stability of stress clashes (e.g. *cò-áuthor*, *dèfórest*, *èx-áctor*), the fact that vowel-final prefixes are near-systematically realized with diphthongs (e.g. /èj/*political*, /dɪj/*stabilize*) and the possibility to find geminated consonants at the prefix-word boundary (e.g. *i[m:]oral*, *i[l:]egal*). Those observations are consistent with an analysis in which those prefixes have their own phonological word which dominates a foot, which itself must be at least bimoraic: *dèfórest* (,dɪj)ω('fɔɪst)ω, *àpolitical* (,ɛj)ω(pə'lɪtɪkəl)ω, *immoral* (ɪm)ω('mo:ɪəl)ω.

3.2.2. Stratal Phonology

Stratal Phonology, which has mainly been developed by Paul Kiparsky (2000, 2015) and Ricardo Bermúdez-Otero (2011, 2012, 2018b), is a general theory of how the phonology interacts with the other components of the grammar. It takes over some of its main concepts from the earlier literature, notably the **cycle** from generative phonology (Chomsky *et al.* 1956) and **stratification** mainly from Lexical Phonology (Kaisse & Shaw 1985; Kiparsky 1982, 1985; Mohanan 1982), although the concept has a deeper history (see Bermúdez-Otero 2018b). Stratification is the idea that morphosyntactic constituents are divided into three different categories, stems, words and phrases, which trigger the application of different phonological grammars (it is said that they define “cyclic domains”) as in (30), which is taken from Bermúdez-Otero (2018b: 105).

- (30) a. Roots do not define cyclic domains.
 b. Some stems define cyclic domains for the stem-level phonology (\mathcal{P}_{SL}).
 c. Words define cyclic domains for the word-level phonology (\mathcal{P}_{WL}).
 d. Utterances define cyclic domains for the phrase-level phonology (\mathcal{P}_{PL}).

Unlike rule-based Lexical Phonology, Stratal Phonology adopts the **parallel computation** of OT, and so each stratum has its own constraint hierarchy. Moreover, stem-level phonology applies before word-level phonology, which itself applies before phrase-level phonology, as shown in Figure 10. The stem-level phonology may apply several times but the word-level and phrase-level phonologies apply only once.

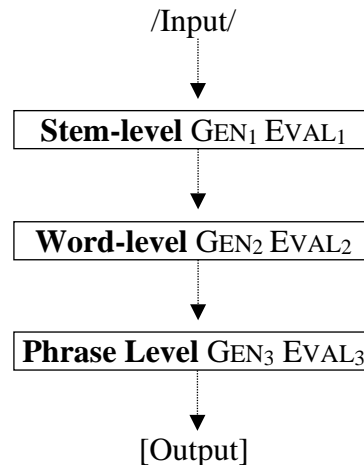


Figure 10. Stratal Phonology grammar

The serialism of a model such as Stratal Phonology can capture cases of morphosyntactically induced phonological opacity, which can be defined as in (31).

(31) **Definition of phonological opacity** (after Bermúdez-Otero 1999: 58)

A rule \mathcal{R} of the form $A \rightarrow B / X_Y$ is opaque to the extent that there are surface representations in the language having

- either (a) A in the environment X_Y (underapplication)
- or (b) B derived by \mathcal{R} in an environment other than X_Y (overapplication)

Because of the adoption of the parallelism of OT on each stratum, opacity may only arise through the successive steps of the computation (e.g. between different stem-level cycles, or cycles of different strata) and so should be subject to Cyclic Containment:

(32) **Cyclic Containment** (Bermúdez-Otero 2018b: 103)

In cases of morphosyntactically induced phonological opacity, a linguistic expression inherits its opaque phonological properties from a constituent defining an immediate cyclic subdomain.

This predicts opacity to be considerably more restricted than in rule-based Lexical Phonology, which allows for intrastratum opacity to arise through rule ordering (a recent example of this is Jensen (2022)). One of the central predictions of the theory is exposed in (33).

(33) **The Russian Doll Theorem** (Bermúdez-Otero 2011)

Let there be the nested cyclic domains $[\gamma \dots [\beta \dots [\alpha \dots] \dots] \dots]$. If a phonological process p is opaque in β because its domain is α , then p is opaque in γ .

Corollary: If a phonological process exhibits cyclic misapplication within a certain phonological configuration created by affixation, then it must also exhibit cyclic misapplication if the same configuration arises by word concatenation.

An illustration of this can be found in the incidence of /l/-darkening in different varieties of English as exposed in Bermúdez-Otero (2011: §9; see references therein). As can be seen in (34), the stratal affiliation of /l/-darkening, like most phonological processes, sees its domain become narrower over time, ‘climbing’ from one stratum to the next. This results in a Russian Doll pattern such that whenever a prevocalic dark [ɫ] which is a case of morphosyntactically induced overapplication can be found in the morphosyntactic environment of a given stratum, it is also found in that of the lower stratum. For example, in the “Am1” variety (described by Sproat & Fujimura (1993)), the fact that /l/ darkens in *heal it* but not in *healing* suggests that /l/-darkening is a word-level process as /l/ is in the coda at the word level in the former but not in the latter. If the process applies at the word-level, the theory predicts that an /l/ that was in the environment for darkening at the word-level will remain dark at the phrase-level, be it in the coda (e.g. *heal ɫ*) or in the onset (e.g. *heal it*, with resyllabification). There are no cases in which /l/-darkening would apply in *heal* and *healing*, but not in *heal it*.

(34)

	RP	Am1	Am2
<i>Healey</i>	l	ɫ	ɫ
<i>heal-ing</i>	l	l	ɫ
<i>heal it</i>	l	ɫ	ɫ
<i>heal ɫ</i>	ɫ	ɫ	ɫ
darkening of rhyml /l/ applies at...	PL	WL	SL

An important aspect of the theory is that it assumes, following Jackendoff (1975), that complex words can be listed in the mental lexicon, and that there can be different types of lexical listing. It is postulated that the outputs of the stem-level phonology are **stored “nonanalytically”**, i.e. as undecomposed forms, while the word-level and phrase-level constructs are either **stored “analytically”** or **not stored at all**. Crucially, what is nonanalytically stored is the phonological representation, and other levels of representation (semantic, morphosyntactic) may retain internal complexity. For example, an idiom such as *to pull X’s leg* must be stored because its semantic representation is non-compositional, but its morphological and phonological representation cannot be a single whole-form representation because the expression can be affected by processes such as wh-movement (e.g. *Whose leg did you pull?*; Bermúdez-Otero 2013). A consequence of nonanalytic listing is that stem-level phonological processes are assumed to function like “lexical redundancy rules”, which means

that they capture the regularities found in stored items but may also be used ‘on-line’ to generate novel forms.

Nonanalytic listing is one of the components that the theory uses to capture effects that used to be analysed as effects of stratum-internal cyclicity in previous cyclic models which assumed that the stem-level was internally cyclic (e.g. Lexical Phonology). Another component that is necessary to capture such effects is morphological blocking⁵¹, i.e. the assumption that the existence of a stored complex form will ‘block’ the on-line computation of the complex form from its parts. Along with high-ranking faithfulness at the stem-level, those components can capture the fact that stem-level processes may sustain lexical exceptions. For example, *Árabic* has antepenultimate stress, unlike the vast majority of adjectives in *-ic*. It is assumed to stay that way because its representation, /'aɪəbɪk/, is stored nonanalytically, because morphological blocking prevents the derivation from *Árab* from generating */ə'ɪabɪk/ (cp. *átom* ♦ *atómic*), and because a high-ranking faithfulness constraint enforces the preservation of the stored stress. The same mechanisms can account for “Weak Preservation” (i.e. the preservation of a primary stress of the base as secondary stress in the derivative; the term is from Burzio 1994), which is analysed as a ‘cyclic’ effect: in a word like *Elizabéthan*, the default rules would predict initial secondary stress (see §2.3). Under the assumption that a language user has to compute that word on-line, but has a nonanalytically stored entry for *Elizabeth*, blocking will prevent derivation from the root *Elizabeth*, and high-ranking faithfulness will ensure that the stress in *Elizabeth* is preserved in *Elizabéthan*. The fact that the same mechanisms are used to capture lexical exceptions and cyclic effects predicts that a connection will always exist between the two. This is formulated as “Chung’s Generalization”, shown in (35).

(35) **Chung’s Generalization** (from Bermúdez-Otero 2012 and after Chung 1983: 63)

If a stem-level phonological process can sustain lexical exceptions in monomorphemic items, then it can show cyclic reapplication in complex stem-level forms, and vice versa.

Thus, the existence of complex derivatives with second-syllable secondary stress such as *Elizabéthan* predicts that there should be lexical exceptions among monomorphemic (or deradical) words with the same stress pattern, and vice versa. As was seen in §2.3, this is indeed the case (e.g. *Epàminóndas*, *egàlitárian*).

Let us conclude this section with how the theory attempts to tease apart representational effects from prosodic effects. Proposals on this point can be found in Bermúdez-Otero & Luís (2009). Four criteria are proposed:

- **Phonetics:** as prosodic structure is phonetically interpretable but cyclic domains are not, we should see phonetic effects of the former but not the latter. For example, they report (based on previous literature) that there is no preboundary lengthening of /tj/ in the pseudowords *beel-ic* and *beel-ing*, which suggests that the phonological difference between the two classes of English suffixes is stratal and not prosodic, *contra* proposals according to which ‘neutral’ affixes are prosodified outside of the phonological word of

⁵¹ Note that morphological blocking is assumed to be sensitive to lexical frequency, so that a highly frequent derivative relative to its base may block the computation of the phonological representation of the derivative from that base, while a highly frequent base relative to its derivative will be less likely to be blocked. Those assumptions are taken over from dual-route race models, which will be discussed in more detail in §3.7.

the base: $((beel)_\omega ing)_\omega$, which would incorrectly predict preboundary lengthening of /ɪj/.

- **Variation:** a variable phonological process should have identical application rates in structures with identical cyclic or prosodic structures. For example, Hayes (2000) reports that word-level constructions (e.g. *healing*) show rates of /l/-darkening that are lower than phrase-level constructions (e.g. *heal it*) and higher than stem-level constructions (e.g. *Healey*). There are different analyses of the prosodification of word-level constructions, which would either predict that they should pattern with stem-level or phrase-level constructions, incorrectly in any case. Bermúdez-Otero & Luís thus argue that the effect is therefore stratal and not prosodic.
- **Bracket Erasure:** in this model, phonology can only access the morphosyntactic information of the current cycle, but not that of earlier cycles. Therefore, a phrase-level process such as flapping should not be influenced by the morphosyntactic structure of words as this structure is not available at the phrase-level. If differences arise between derivatives (e.g. *câpi[r]alistic* vs. *mîli[t^h]aristic*), it is argued that they must reflect a prosodic difference built at earlier levels that is retained up to the phrase-level.
- **Coextensiveness of morphosyntactic categories and cyclic domains:** Bermúdez-Otero & Luís provide a detailed discussion of the phonological behaviour of proclitics and enclitics in European Portuguese and show that both cyclic domains and prosodic units are necessary, and that a key difference between the two is that prosodic units need not be coextensive with some morphosyntactic category, but that cyclic domains do.

Many additional aspects of the theory could be discussed, but those I have just reviewed are those that will be the most relevant for the following sections.

3.2.3. Output-Output Correspondence

One of the theoretical innovations of OT was the fact that phonological generalizations are enforced through constraints that are **evaluated in parallel**, as opposed to the serialism that characterized earlier generative approaches since *SPE*. OO-correspondence (Benua 1997; Burzio 1996; Kenstowicz 1996) is an extension of Correspondence Theory, which was put forward by McCarthy & Prince (1995, 1999). McCarthy & Prince introduced MAX, DEP and IDENT constraints, which mapped the input with the output (IO-correspondence). OO-correspondence constraints (also known as Transderivational Correspondence, Paradigm Uniformity or Anti-Allomorphy constraints) were introduced to model paradigmatic dependencies between words while enforcing strict parallelism. Those two types of correspondence relationships are represented in Figure 11.

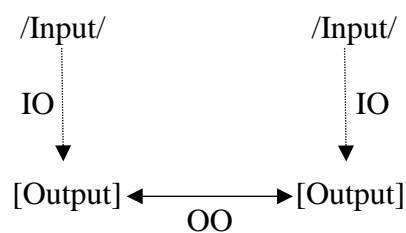


Figure 11. OO-correspondence (after Collie 2007: 232)

While this model can be used to capture local paradigmatic dependencies, i.e. between a derivative and its local base, it can be used to capture dependencies that are not strictly local, and in which a derivative appears to be faithful to a member of its morphological family which is not its local base. This type of model has been used recently with cases of “lexical conservatism”, which will be discussed in §3.5.6.

Regarding the division of labour between representational and procedural morphosyntactic conditioning, Raffelsiefen (2005) makes proposals in that framework to tease apart what she calls “**Paradigm Uniformity**” effects (= OO-Correspondence) and “**boundary effects**” (= prosodic effects). She argues that Paradigm Uniformity should not be invoked if the processes can be accounted for by regular boundary effects. Such boundary effects notably include the effects of syllabification: fortition of consonants in onset position, lenition in coda positions and lengthening of the nucleus in open syllables. She assumes a strict version of the Strict Layer Hypothesis, so that prosodic units should always be strictly contained in the superordinate prosodic unit of which it is a part. Thus, she assumes that consonant-initial suffixes are “non-cohering” in that they do not syllabify with the base that they are attached to. She assumes that the regular syllabification of a consonant that is found between a stressed vowel and an unstressed vowel is for this consonant to be ambisyllabic, which means that it will be lenited. However, she cites phonetic evidence that this is not the case of the initial consonant of suffixes such as *-ness*, so that the /n/ in *shyness* is stronger than that of *minus*, arguably because of different prosodifications: (*shy*)_o*ness* vs. (*minus*)_o. Therefore, she argues that this can explain the stress neutrality of such suffixes (e.g. as shown by preantepenultimate stress in *radicalness*) and that one need not invoke Paradigm Uniformity because the boundary effect, which is independently needed, can also account for stress neutrality: the suffix is not in the domain of stress assignment. More broadly, she argues that all “identity effects” which result from non-cohesion should not be treated as effects of Paradigm Uniformity and discusses a number of processes found in compounds, prefixed words and suffixed words. More details of her analysis of the different behaviours of suffixed words will be seen in §3.5.

OO-correspondence has been undergoing severe criticism from the defenders of cyclic models, which I have summarized in my dissertation (Dabouis 2016b: §4.2.2.2) and reproduce here:

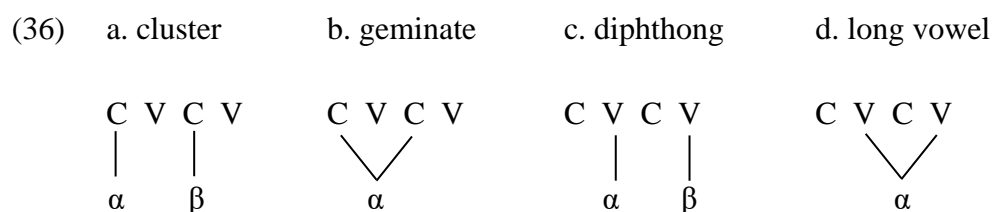
- While Benua (1997) only assumed that potential bases were only those members of the morphological family which are separated from the derivative by a single affix, definitions of the term ‘base’ have become increasingly vague, and it is not always clear which words may act as phonological bases and on what grounds.
- OO-correspondence predicts that we may have derivative-to-base paradigmatic dependencies, although there is hardly any evidence for such dependencies (see Bermúdez-Otero 1999: §3.4.1).

- The analysis of morphosyntactically induced misapplication (i.e. phonological opacity) requires the existence of a base in which the process applies transparently and which can then transfer its properties onto a word in which it is opaque. Bermúdez-Otero (2018b: 119) reports seven instances of morphosyntactically induced misapplication for which no related surface form is transparent.

3.2.4. *Strict CV*

Strict CV is a version of **Government Phonology**⁵² which was initially developed by Lowenstamm (1996, 1999) and Scheer (2004). Government Phonology has its origins in autosegmental phonology, and it initially focused on the **internal structure of segments**, which gave rise to Element Theory (see e.g. Backley 2011). That theory was a response to some of the issues with feature-based theories, mainly (according to Scheer & Kula (2018)) overgeneration (the more features a model has, the larger the set of possible segments) and the lack of inherent natural class predictiveness (not all features may be used to identify natural classes). Unlike features, elements are unary and are phonetically interpretable (e.g. you cannot pronounce [+labial] but you can pronounce |U|).

The second area of focus of Government phonology has been **syllable structure** (see the overview of Scheer & Cyran (2018b)). The core of its research program has been to replace arboreal structures by lateral relations between segments to capture generalizations about “syllable” structure. The units that can be engaged in such lateral relations include empty skeletal positions (C or V). This is particularly common in Strict CV, in which it is assumed that CV is the only possible syllable type. Thus, clusters, geminates, diphthongs and long vowels are assumed to have the representations in (36), which are taken from Scheer & Cyran (2018b).

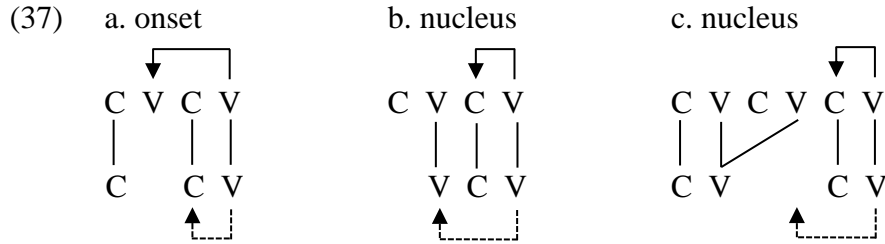


In (36), the skeletal plane is only composed of C and V slots, and those are associated to melodic entities (α and β) which stand on a different autosegmental plane. What can be seen is that skeletal positions may be empty, i.e. they are not connected to melodic material.

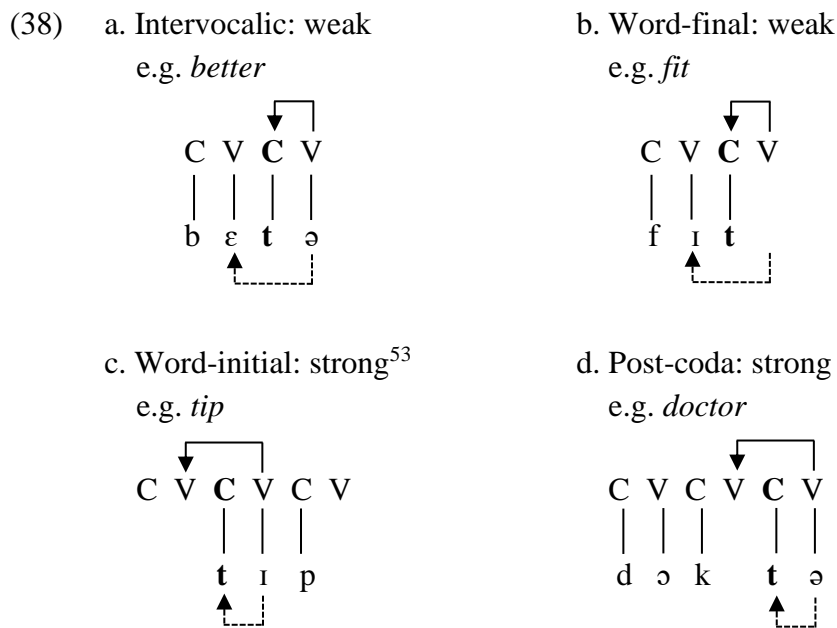
There are various models of the theory with different inventories of lateral relations but, in Strict CV and especially the version called the Lateral Theory of Phonology (Scheer 2004), there are only two lateral forces: Government and Licensing. Government is assumed to be an inhibitory force while Licensing has a strengthening effect. Both relations are assumed to be right-headed, i.e. they go from right to left, and they are by default headed by melodically filled nuclei (i.e. vowels). In certain languages, final empty nuclei may also be lateral actors. As a default, a nucleus governs the preceding V position if it is empty, otherwise it governs the onset that

⁵² See Scheer & Kula (2018) for a review of the historical developments of Government Phonology.

precedes it. Licensing targets the preceding position that has escaped Government, as illustrated in (37), where Government is represented by a solid line and Licensing by a dotted line.



If Government targets a preceding empty nucleus, then Licensing will target the preceding onset (37a), but if Government targets the preceding onset (because the preceding nucleus is filled), then Licensing will target the preceding nucleus (37b, c). Those lateral forces may be used to capture the distribution of strong and weak consonantal positions (see e.g. Ségéral & Scheer 2001, 2008). For example, the distribution of the allophones of /t/ in English may be formalized as in (38).



In (38a, b), /t/ is governed by the nucleus that follows it⁵⁴ and is thus weak, which means that it will be realized with one of the weak allophones of /t/, depending on the variety of English (e.g. [t^ˈ], [r], [ʔ]). However, in (38c, d), the vowel governs the preceding empty nucleus and so /t/ undergoes Licensing and is thus strong, which means that it will be realized with a strong allophone of /t/, usually [t^h].⁵⁵

⁵³ In English, words are assumed to begin with an empty CV unit. It is represented only here as this is the only of those examples in which it is relevant. Details on this assumption are presented below.

⁵⁴ It is assumed that the final empty nucleus is a lateral actor in English, based on word-final phonotactics: the existence of words such as *bend* or *act* is taken to be an indicator that it can govern the empty nucleus that intervenes between the two final consonants (Scheer & Cyran 2018b: 277).

⁵⁵ Ségéral & Scheer (2008) note that in the post-coda position, there may be /t/-lenition if the preceding consonant is a sonorant (e.g. *quarter*, *winter*). Under the assumption that diphthongs are vowel + glide sequences, we might

Another important claim in the theory is that, in some languages, words begin with **an initial empty CV unit** (Lowenstamm 1999). This has been used to capture different generalizations that can be observed at the left edge of words (see Table 10) and may be tied together so that, if a language has one of them, it should have the other ones, as all are predicted by the presence or absence of the initial CV (Ségéral & Scheer 2008, Scheer 2014).

	Initial CV present (e.g. English, German, Italian, French ⁵⁶)	Initial CV absent (e.g. Moroccan Arabic, Polish, Russian, Greek)
Word-initial clusters	TR only	Anything (TR, RT, RR, TT)
First vowel of words may alternate with zero	No	Yes
Word-initial consonants are strong	Yes	No

Table 10. Predictions made by the presence or absence of the initial CV (T stands for an obstruent and R for a liquid)

The presence of this empty CV makes predictions based on the usual workings of Government and Licensing. For example, it can be used to account for the strength of initial consonants because the nucleus of the initial empty CV needs to be governed, and so Licensing may target the initial onset (see (38c)), thus strengthening it.

Now that we have seen some of the basic components of the theory, let us turn to how it incorporates morphosyntactic information. Like Stratal Phonology, Strict CV, and especially the Lateral Theory of Phonology, is very explicit on how it deals with interfaces, within a strongly modular framework. The procedural device used to incorporate morphosyntactic information into the phonology is, like Stratal Phonology⁵⁷, cycles, although they are called “**phases**” because the view of morphosyntax adopted is that of Phase Theory (Chomsky 2000). However, not all syntactic phases seem to have phonological manifestations, but when they do, it is assumed to be in the form of an initial CV. Therefore, the initial CV presented above is not word-initial, but phase-initial (Scheer 2009, 2012b: §307). Thus, inside-out interpretation from smaller to bigger phonological strings through phases is the procedural device used to carry morphosyntactic information into the phonology, and the insertion of initial CV units is the representational device.

expect them to pattern with closed syllables, as they would have the same structure as (38d). In General American, we may have flapping after a diphthong just as after a vowel, but also after /ɪ/. There are varieties in which that is not the case: in the New Zealand Acrolect, diphthongs (and long vowels) pattern with closed syllables (Bye & de Lacy 2008). Szigetvári (to appear) assumes that sonority is what conditions flapping, and that different varieties may pick a different point in the sonority scale regarding which consonants condition flapping. Thus, it appears that Government and Licensing cannot fully explain that process and that other tools are required.

⁵⁶ Although French is undergoing change and may be losing its initial CV according to Pagliano (2003), as evidenced by the loss of first-syllable schwas that may lead to initial RT clusters (e.g. [vəpχã] → [χpχã] *reprends* ‘Take again 2SG-PRES-IMP’).

⁵⁷ A big difference with Stratal Phonology is that it is assumed that there is only one phonological grammar, and so it is assumed to apply at different levels in morphosyntactic structure.

The use of syllabic space to carry boundary information into the phonology is defended by arguing that it is not a diacritic (see e.g. Scheer & Cyran 2018a: 301-302). As the theory has clear rules on how C and V slots should interact, inserting an empty CV unit on the left edge of a string makes specific predictions as to what kinds of phonological processes may be found in that position. This is opposed to other representational objects, mainly the Prosodic Hierarchy, which are diacritics: prosodic constituents mark morphosyntactic boundaries but make no specific predictions as to the nature of phonological processes that one should expect to find at their edges. Thus, defendants of the Strict CV approach to the interface often formulate harsh criticism against the Prosodic Hierarchy (see e.g. Newell 2023; Scheer 2008, 2012a).

3.3. Semantically opaque morphology and phonology

A common assumption in the Guierrian School is that there is a class of historically prefixed words that have no clearly identifiable meaning (e.g. *accept*, *construct*, *decide*, *submit*) and yet have phonological behaviours that set them apart from truly simplex words (Dabouis *et al.* 2023: §2.2.1). When I was trained in this framework, I accepted this as a fact of English morphophonology but, when I exchanged with researchers from other research traditions, this ‘fact’ was often questioned or rejected altogether. I was brought to look deeper into the characteristics of those words whose morphological structure has somehow become opacified. I looked into this with Jean-Michel Fournier as we were studying primary stress placement in verbs (see §3.4.2), and on my own in my dissertation and for a presentation at the 14th edition of the Old World Conference in Phonology (OCP; Dabouis 2017). Eventually, we wrote down a paper that brings together all the elements that we had gathered (Dabouis & J.-M. Fournier 2024). The research that I have done on this issue has mainly consisted in gathering new evidence on the phonological manifestations of such opaque morphological structures, replicating existing results and reviewing the available evidence from various sources in phonology, morphosyntax and psycholinguistics. The main elements that have come out of this research and which are gathered in the joint paper with Jean-Michel Fournier are presented in §3.3.1. Subsequently, I also studied a category of words which display comparable characteristics: historically complex proper names (e.g. *Cambridge*, *Oxford*, *Washington*). I gave two presentations on the issue, at the 61st congress of the SAES (Société des Anglicistes de l’Enseignement Supérieur; Dabouis 2022) and at the 20th OCP conference (Dabouis 2023b). The results of this second investigation are presented in §3.3.2. We have also worked on the issue of learnability, which will be discussed in §3.3.3. In both papers, possible representations of the lexical entries for those two types of words and the generalizations that might be learned about those words are proposed, and will be discussed in §3.3.4.

3.3.1. Historically prefixed words

Within the Guierrian School, the existence of different types of prefixed words is accepted by all. Guierre (1979) defined two classes of prefixes which he called “**separable**” and “**inseparable**”, which were distinguished by their semantic transparency. In later works, it has

been widely acknowledged that this is a simplificatory view and that “separability” should be seen as:

- a property of a prefixed word rather than that of a prefix as certain prefixes can be found in words which vary in semantic transparency (Abasq 2009; Videau 2013);
- a gradient rather than a categorical distinction (Guyot-Talbot 2003; Videau 2013) as certain types of prefixed words do not neatly fit into one of the two categories. This gradience may depend on a number of characteristics discussed by J.-M. Fournier (1996), though not in those terms.

In *SPE*, a similar distinction has been made between prefixes that are introduced by the boundary =, found in “complex verbs” which are “morphologically analysable into one of the prefixes *trans-*, *per-*, *con-*, etc. followed by a stem such as *-fer*, *-mit*, *-cede*, *-curn* or *-pel*” (*SPE*: 94) and prefixes introduced by the boundary #, which they hardly discuss, but roughly correspond to productive, semantically transparent prefixes (see the oppositions between *re=solve* ‘determine’ vs. *re#solve* ‘solve anew’ or *re=serve* ‘withhold’ vs. *re#serve* ‘serve anew’ p. 95).⁵⁸ In Dabouis & J.-M. Fournier (2023), we trace how prefixed words that were analysed as having an = boundary have been increasingly analysed as simplex words.

Over the years, we have been recurrently faced with criticism from researchers from various theoretical frameworks who were rejecting the idea that what Guierre called inseparable prefixes were prefixes. Thus, we took the criticism seriously and tested their implicit claim: if words such as *admit*, *deceive*, *retain* or *subtract* have no internal structure at all, we should expect them to pattern with simplex words. The following sections deal with that question by reviewing the main pieces of available evidence that inform us on that claim. As will be seen, the vast majority of the evidence points in the direction of internal complexity.

As I have mentioned above, there is definitely gradience in the extent to which prefixed words are transparent (see §3.3.1.4), and probably as much diversity in how language users actually perceive them. If we see the decomposability of prefixed words as a continuum⁵⁹, it is to be expected that the least decomposable ones will be not perceived as complex by as many language users as the most decomposable ones. However, given the results that will be reviewed in the following sections, it would appear that there are sufficient cues for them to be perceived as complex by a sufficient number of language users so that their overall behaviour is distinct from that of simplex words. As there are no established procedures for identifying such opacified morphological structures synchronically, most of the studies reported in the following section identify words as prefixed based on etymology (using the *OED*) and exclude items in which the meaning is (near-)compositional (e.g. *co-author*, *deconstruct*, *rearrange*) and those for which the prefix appears to bear a meaning that corresponds to the one it usually has in compositional constructions, even though it is attached to a bound base (e.g. *decelerate*, *decrease*, *dissimilate*).

⁵⁸ *SPE* also uses the + boundary for “prefixed” words which are in fact all neoclassical compounds (e.g. *para+site*, *chromo+somal*, *philo+sophical*; *SPE*: 95).

⁵⁹ On theoretical grounds, J.-M. Fournier (p.c.) rejects the idea of a continuum in favour of gradience, which would be influenced by various factors (e.g. the semantics of the base, of the prefix, the recurrence of elements).

3.3.1.1. *Phonological evidence*

As the Guierrian School has long assumed that opaque prefixation plays a role on stress assignment, it has been tested in a number of studies using data from pronunciation dictionaries. The broadest one is Guierre (1979), in which he reports distinct stress patterns for disyllabic non-prefixed and (inseparable) prefixed words, as shown in Figure 12. Note that the difference between prefixed and non-prefixed words is statistically significant in all syntactic categories.⁶⁰

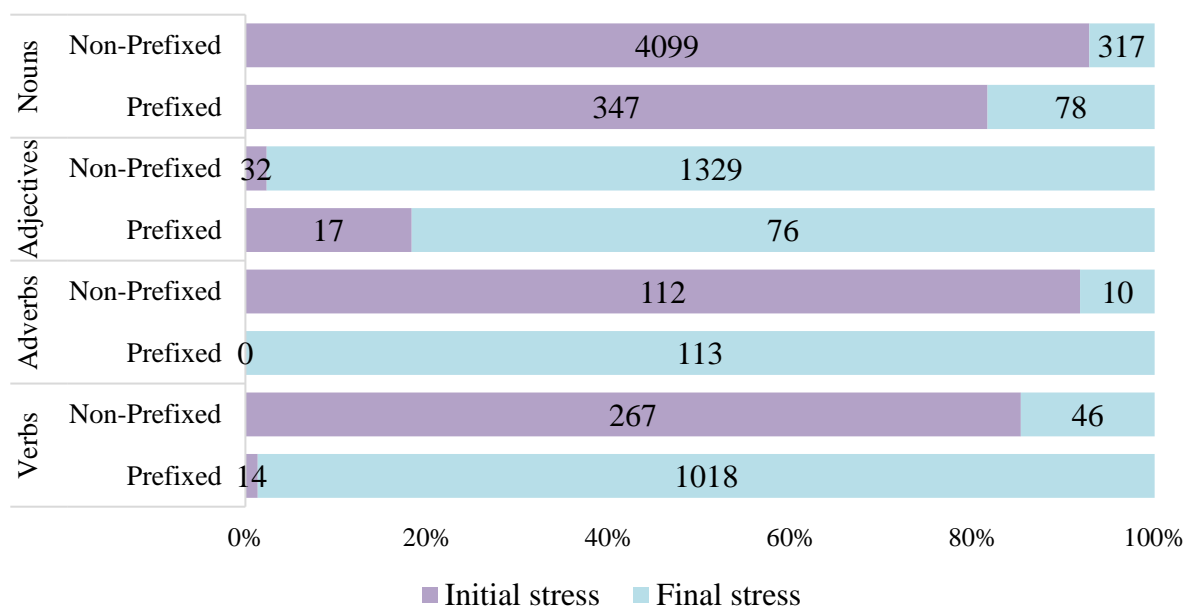


Figure 12. Position of primary stress in disyllabic nouns, adjectives, adverbs and verbs according to Guierre (1979)

The results for verbs has been replicated while controlling for all other morphological influences and for syllable weight in Dabouis & Fournier (2023; see §3.4.2) using more recent dictionary data. In that paper, we also replicate Halle & Keyser's (1971: 68-69) and Guierre's (197: §1.2.2) finding that, in American English, there is a significant difference regarding the position of primary stress in disyllabic verbs in <-ate> depending on whether or not they are prefixed. Prefixed verbs in <-ate> such as those in (39a) have final primary stress while non-prefixed ones such as those in (39b) have initial primary stress.⁶¹

⁶⁰ Nouns ($\chi^2 = 63.119$, $df = 1$, $p < .001$), adjectives ($\chi^2 = 786.02$, $df = 1$, $p < .001$), adverbs ($\chi^2 = 194.53$, $df = 1$, $p < .001$) and verbs ($\chi^2 = 1018.9$, $df = 1$, $p < .001$). This was not done by Guierre, but done in Dabouis & J.-M. Fournier (2024).

⁶¹ Halle & Keyser and J.-M. Fournier (1996) treat words such as those in in (39a) as non-suffixed and those in (39b) as suffixed, but it is not always easy to see on what this distinction was based as the bases are not always found in other words.

- | | | | | |
|------|----|---|----|--|
| (39) | a. | abláte
colláte
debáte
infláte
reláte
transláte | b. | cástrate
díctate
grádate
mígrate
rótate
víbrate |
|------|----|---|----|--|

In longer verbs, *SPE* proposed a rule called the Alternating Stress Rule (ASR) that is formulated as in (40).

(40) **Alternating Stress Rule**

$V \rightarrow [1 \text{ stress}] / _ _ C_0 (=) C_0VC_0 [1 \text{ stress}] C_0]_{NAV}$

This rule states that primary stress is moved two syllables leftwards if it has been assigned to a final syllable, and there may be an = boundary between the antepenult and the penult (e.g. *de=tonate*, *re=plicate*). However, as rules may be blocked by =, if such a boundary appears between the penult and the final syllable (e.g. *contra=dict*, *inter=sect*), the ASR will be blocked. Siegel (1974) made a different claim: she claimed that the ASR only applies if the final syllable is a suffix and dropped the reference to =. As will be seen in §3.4.2, the available evidence rather suggests that *SPE*'s analysis was the correct one, as indeed words with two prefix syllables usually have final primary stress, while other words have antepenultimate primary stress if they have a final heavy syllable, even if they are not suffixed.

As will be seen in §3.7.2.2, Sabine Arndt-Lappe and I have also found evidence that opaque prefixation interacts with Weak Preservation, so that monosyllabic opaque prefixes favour second-syllable stress preservation. Therefore, such opaque prefixes seem to “repel” primary stress in words that are not nouns, to prefer to stay unstressed in derived words, and there has been a longstanding observation in the literature that they also usually have unstressed/reduced vowels in the initial pretonic position. Indeed, most major works on English make that observation as it distinguishes prefixed words from non-prefixed words, especially in initial closed syllables (e.g. */a/mpóon*, *p/ɔ/ntificate* cp. */ə/dvámce*, *c/ə/ndénse*; see §2.2.2; *SPE*: 118; Collie 2007: 129, 215, 318-319; Guierre 1979: 253; Halle & Keyser 1971: 37; Halle & Vergnaud 1987: 239; Hammond 2003; Hayes 1982; Liberman & Prince 1977; Pater 2000; Selkirk 1980b).⁶² Thus, in the study of vowel reduction that I have started reporting in §2.2.2, we sought to quantify this generalization and evaluate whether or not it holds alongside other predictors of vowel reduction. We indeed found a clearly marked difference between opaque prefixed words and non-prefixed words in all subsets of the data, and the effect holds in the regression analyses that we conducted. Figure 13 shows a simplified representation of the distribution of the data for the different categories of words that were investigated in the paper, taking only into account the main pronunciation given in Wells (2008).

⁶² Note that this may be true of SSBE and General American but not of northern varieties of England, where a number of prefixes in closed syllables maintain full vowels (e.g. *conclude* /kɔŋ'klúwd/, *advance* /ad'vans/; see Cruttenden 2014: 139).

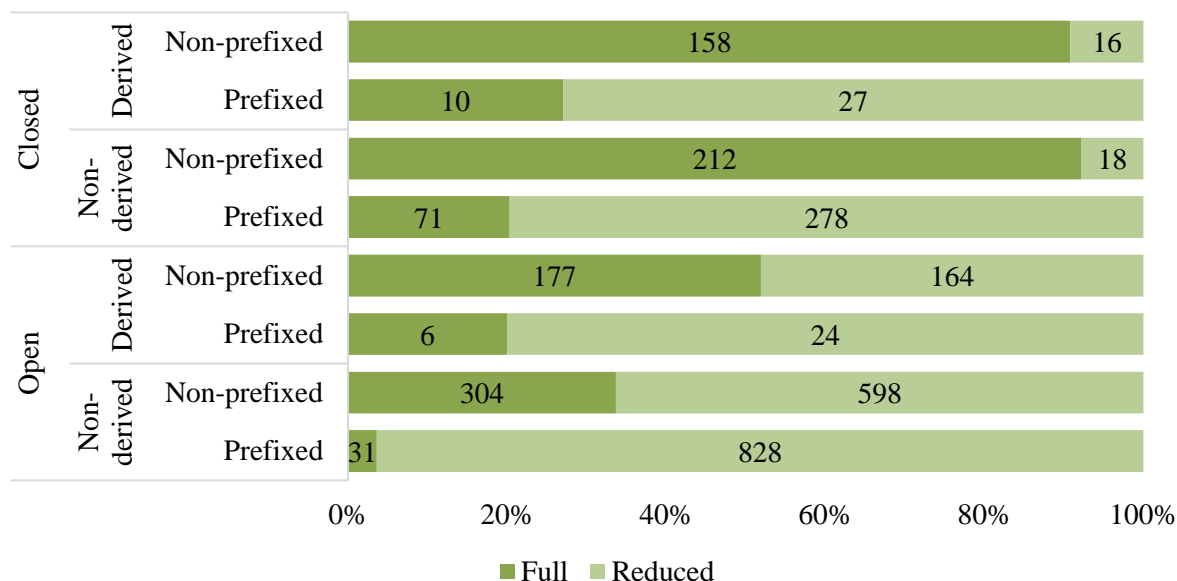


Figure 13. Nature of vowels found in the initial pretonic syllable of prefixed and non-prefixed words that are stress-shifted derivatives (‘derived’) or monomorphemic words or words formed of a bound root and a suffix (‘non-derived’), in open or closed syllables (based on Dabouis & J.-M. Fournier to appear)

As can be seen from Figure 13, prefixed words systematically have reduced vowels more often than non-prefixed words. The difference is greatest in closed syllables, but it is still clearly present in open syllables too.⁶³

Another phonological behaviour that sets opaque prefixed words aside from simplex words is also related to vowel reduction. Indeed, Guierre (1979: §4.2.6) observed that non-prefixed words generally have reduced vowels in the second syllable of disyllabic words with initial primary stress in 87% of cases (see (41a)) while “inseparable” prefixed words have full vowels in 84% of cases (see (41b)).

- | | | | | |
|------|----|-----------------------|----|-------------|
| (41) | a. | álb/ə/rt | b. | ádv/ə:/rt |
| | | báll/ə/st | | díg/ε/st |
| | | hárv/ɪ/st | | cóntr/a:/st |
| | | hón/ɪj/ ⁶⁴ | | súrv/εj/ |
| | | sérp/ə/nt | | cómm/ε/nt |
| | | pílf/ə/r | | tránsf/ə:/r |

However, there is a possible confound for this difference: the words in (41b) are often deverbal nouns. It has been noted that disyllabic nouns with initial primary stress that have a related verb with final stress (e.g. *escort*, *import*, *project*) tend to have full vowels in their second syllable (Fudge 1984: 32, 167; Poldauf 1984: 38).⁶⁵ There is a considerable overlap between this factor

⁶³ We also found that opaque prefixed words were clearly different from transparent prefixed words, the latter almost systematically having a full vowel in their first syllable, even if the dictionary does not mark it as “stressed”.

⁶⁴ Guierre analyzed final /ɪj/, transcribed as /i/, as reduced.

⁶⁵ This is viewed as a consequence of Level I cyclicity in Lexical Phonology (Kiparsky 1982), as verb-to-noun “zero derivation” is assumed to take place on Level I, and so may entail stress shifts (e.g. *tormént_V* → *tormènt_N*).

and prefixation, as such deverbal nouns are often prefixed (although there are a few counterexamples such as *férm/ɛ/nt* or *tórm/ɛ/nt*). The only study that has sought to disentangle the two factors is Dahak (2011). She conducted a study using 9,500 disyllabic words with initial primary stress from pronunciation dictionaries. Among those, 5.5% are prefixed, and the proportion of full vowels in their second syllable that she found is very similar to that reported by Guierre (1979). However, she finds that only a quarter of those prefixed words have a related form with final stress, and that they differ from those which do not, as shown in Figure 14.

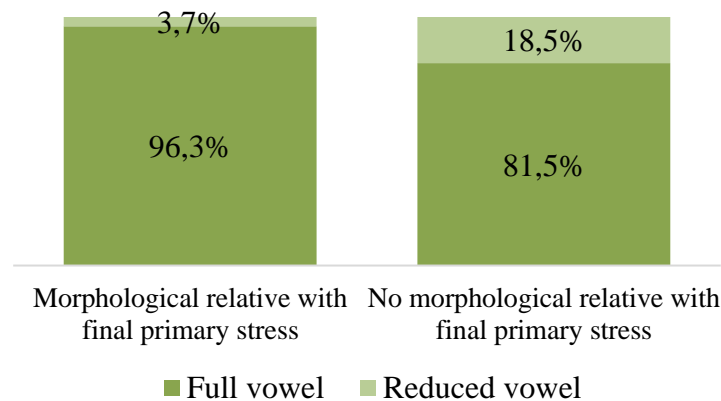


Figure 14. Second vowel in disyllabic nouns with initial primary stress

Thus, although her results are not always detailed and she did not conduct multifactorial statistical analyses, they suggest that the higher proportion of words with full vowels in prefixed words such as those in (41b) cannot be solely reduced to a form of “stress preservation” from the putative verbal bases.

This observation is further supported by a study that I contributed to and which was presented to the 2012 MFM (Abasq *et al.* 2012), which will be discussed in further detail in §3.4.4. In that study, we looked at how disyllabic prefixed noun-verb pairs were stressed in three different pronunciation dictionaries and three varieties of English (British, American, Australian). One of the results that we reported is that the existence of a related form (noun, verb, or stress variant) with primary stress on a syllable disfavoured the presence of full vowels in related words in which the corresponding syllable does not have primary stress. Moreover, we found considerably lower proportions of full vowels in syllables that correspond to prefixes than in those that correspond to roots. That observation cannot be reduced to the fact that those are different positions (initial vs. final), as existing studies do not show that vowel reduction in the initial pretonic position is more common than reduction in the second syllable of a disyllabic word with initial primary stress. This can be accounted for if the two syllables are indeed distinct types of morphological units, but not if they are both part of a single formative. However, at the time I was not as aware of the literature on vowel reduction as I am now, and we did not conduct multifactorial statistical analyses to control for all possible intervening factors. Therefore, it may be interesting to replicate this study so as to get clearer evidence of both the

while noun-to-verb zero derivation is assumed to take place on Level II and so no stress shifts are expected (e.g. *páttérn_N* → **pattérn_V*).

influence of a related form with primary stress on the corresponding syllable and of the difference between the two syllables of those verb-noun pairs.

Finally, the last piece of phonological evidence that points to the presence of a morphological boundary in opaque prefixed words is medial consonant clusters. Guierre (1990) reports that certain clusters which are illicit in simplex words are found in prefixed words and could thus indicate the presence of a morphological boundary. He notes that the only clusters of two obstruents that are attested in simplex words word-medially are clusters of two voiceless obstruents. More precise evidence is provided by Hammond (1999: §3.3.), who has conducted a more extensive survey of distributional regularities in English. He reports that the following types of word-medial clusters are only (or mostly) found in words with Latinate prefixes:

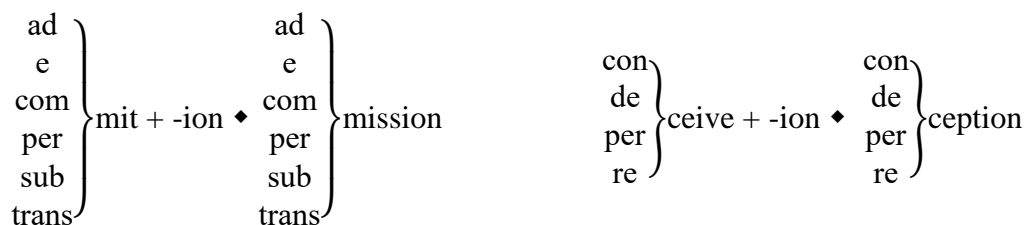
- clusters with voicing disagreements involving voiceless fricatives: /bs/ *abcess*, *absence*, *subsidy*, /bf/ *obfuscate*;
- certain clusters containing a voiced stop followed by a sonorant: /bm/ *submerse*, /dm/ *admire*, /bn/ *obnoxious*;
- most clusters containing a voiced stop followed by a voiced fricative or affricate: /bv/ *obvious*, /bz/ *absolve*, /dv/ *advantage*, /bdʒ/ *object*, /gdʒ/ *suggest*;
- /dh/ in *adhere*, a cluster normally only created by concatenation, e.g. *childhood*, *madhouse*.

As pointed out by an anonymous reviewer who reviewed Dabouis & J.-M. Fournier (2024), it would be interesting to see how those clusters are actually realized phonetically considering that voiced and voiceless plosives may neutralize in preconsonantal position. They suggest that the phonetic realization of such clusters may manifest not through differences in the voicing of the two consonants, but through the duration of the preceding vowel. Because of pre-fortis clipping, we might expect shorter vowels before underlyingly [-voice][-voice] clusters than before [+voice][-voice] clusters.

3.3.1.2. Morphological evidence

Aronoff (1976) argues that bound roots such as *fer* or *mit* are “somehow relevant to the morphology” because the verbs that contain them all have the same standard nominalizing suffix, sometimes along with the same root allomorphy, as shown in (42).

(42)



He also notes that a similar phenomenon can be observed for verbs with the same irregular inflection, even in the absence of semantic similarity (e.g. *get*, *beget*, *forget* – *got*, *begot*, *forgot*

– *gotten, begotten, forgotten*). This analysis could lead one to believe that the roots that alternate in such a way are better cued, and would thus be easier to identify. However, one study sought to test that assumption using a segment shifting study and found no evidence in that direction (Melinger 2001). She presented participants with a source word (e.g. *cooperate, coconut*) in which (pseudo-)affixes are highlighted when the target word is presented (e.g. *author*), and then asked participants to attach the highlighted part to the target word (e.g. *co-author*). The source words could contain free stems, bound roots and words with no internal structure. The words with bound roots contained roots which may or may not have several forms. She found that it was more difficult for participants to extract the (pseudo-)affix in words containing alternating roots than in words containing non-alternating roots. Thus, her study suggests that roots that have a stable form are easier to identify than those that have different forms, but this is the only study available on that issue, and it has a few methodological issues (e.g. roots treated as alternating included cases like *depar/t/* ♦ *depar/tʃ/ure* alongside cases like *revolve* ♦ *revolution*).

3.3.1.3. Psycholinguistic evidence

Although the phonological behaviours displayed by opaque prefixed words regularly differ from those of simplex words, some may be lexicalized, or some might argue that at least some of them are caused by other factors with which opaque prefixation is correlated. Thus, I have been collecting the available evidence from various experimental paradigms in psycholinguistics since 2015 to see if there is independent evidence that the formatives found in such words are detected by language users. In this section, I briefly summarize the available evidence from lexical decision tasks, from studies focused on reading and from ERP studies.

In a **lexical decision task**, participants are presented with a target word, and they are asked to say as quickly as possible whether or not it is a real word. The target may be preceded by a prime which may or may not be related to the target. Reaction times (sometimes called “latencies”) are measured and interpreted in terms of lexical processing. When the prime is related to the target in some way (semantically, phonologically, morphologically), reaction times are normally reduced (this is called “facilitation”). The available evidence from lexical decision task experiments using words or pseudowords on the treatment of prefixed words with bound roots (often called “bound stems” in that literature) such as *conceive, rejuvenate* or *submit* shows that those words are processed differently from simplex words and some studies even suggest that they are processed in the same way as prefixed words with free stems (Coch *et al.* 2020; Taft 1994; Taft *et al.* 1986; Taft & Forster 1975). However, Marslen-Wilson *et al.* (1994) found no facilitation for prefixed words with bound roots as compared to those with free stems is found, but it is possibly because they used cross-modal (visual and auditory) priming, which is known to be particularly sensitive to semantics. Therefore, two studies have used a methodology called masked priming (Forster & Azuma 2000; Pastizzo & Feldman 2004), in which the prime is presented for a much shorter duration and is preceded and followed by a visual mask (usually hashmarks), so that the participants have no awareness of the prime. This methodology is interesting as it neutralizes semantics, and so it is particularly adapted to the study of morphological relationships between words that are semantically opaque. Those two studies found the same amount of facilitation between items sharing a bound root (e.g. *explore* – *implore*) and those sharing a free stem (e.g. *trust* – *distrust*). They also found that the number

of morphological relatives sharing the same root induces further facilitation. This is reminiscent of a result from Taft (1979), who reports greater facilitation for roots which have a higher frequency (e.g. *proach* in *approach* and *reproach* has a higher frequency than *suade* in *persuade* and *dissuade*; lexical decision latencies are reported to be faster for *reproach* than for *dissuade* even though these words have equivalent frequencies). Moreover, semantics may come into play even though prefixed words with bound roots are overall more opaque than those with free stems, as they are not all opaque to the same degree. Thus, Taft & Kougious (2004) report more facilitation for words that contain roots that have a clearer meaning (e.g. *venge* as in *revenge*, *avenge*, *vengeful*, *vengeance*) than those whose roots have a more obscure meaning (e.g. *ceive* in *receive*, *deceive*, *conceive*, *perceive*). Most studies of the lexical decision paradigm have focused on suffixed words and have used visual stimuli. The different available overviews on that rich literature converge on the conclusion that what they call “morpho-orthographic decomposition” is semantically-blind and is applied to all morphologically-structured stimuli in the early stages of recognition (Beyersmann & Grainger 2023; Diependaele *et al.* 2009; Laurie Beth Feldman *et al.* 2009; Heyer & Kornishova 2018; Marslen-Wilson *et al.* 2008; Rastle & Davis 2008). The few studies in spoken-word processing show that a similar mechanism is at work in auditory processing (see the overview in Creemers 2023).

A second type of study that can be found in the psycholinguistic literature are studies which focus on **stress assignment in the reading of disyllabic words**. The foundational study on the issue is Rastle & Coltheart (2000), who sought to establish if stress-irregular words behave similarly to other spelling-to-sound exceptions. The earlier literature had found that exceptions tend to be read slower than words that are regular. For example, orthographic vowel monographs that are followed by two consonants are usually realized as short vowels. Words that conform to this generalization (e.g. *mint* /'mɪnt/) are read faster than words which do not (e.g. *pint* /'pajnt/). They find that this is indeed the case for stress, but only with the following assumptions on “regularity”: disyllabic words have initial stress unless they contain an initial prefix (identified orthographically; e.g. *ad-*, *de-*, *sub-*) or a stress-taking suffix (e.g. *-ette*, *-oon*, *-ade*). They also conduct a production experiment using pseudowords with or without such structures and compared the productions of their participants to the predictions made by a computational model including those factors, and found a very high similarity between them. Ktori *et al.* (2016) conducted a production study on English disyllabic words with patients with acquired surface dyslexia, which is “an acquired disorder of reading in which the reading aloud of irregular words is impaired while the reading aloud of nonwords is spared”. The participants had to read disyllabic prefixed words and non-prefixed words. Prefixed words could be “regular” (if they had final primary stress) or “irregular” (if they had initial primary stress). Their results showed that participants produce significantly more errors on “irregular” prefixed words (e.g. *reflex*) than on the other two types of words (e.g. *remind*, *climate*). Finally, two more recent studies (Ktori *et al.* 2018; Treiman *et al.* 2020) report the results of different nonwords production tasks, in which the presence of a prefix (implemented following Rastle & Coltheart’s (2000) procedure) was tested along other covariates (e.g. syntactic category, number of initial and final consonants) and was found to be a significant predictor of the position of primary stress in the productions made by the participants of those studies. Treiman *et al.* (2020) also report similar results using dictionary data. In those studies, prefixation is thus one of the factors that is probabilistically associated with final stress.

The last type of experimental paradigm that I have reviewed are **event-related potential (ERP)** studies. There are two such studies in which participants were presented with words and nonwords and had to perform a lexical decision task. The authors measured the electrical activity of the brain of the participants, focusing on the N400 (a negative drop occurring at 400ms after the presentation of a stimulus), which is known to be sensitive to lexical status (word vs. nonword) and the frequency of the stimulus. Less frequent words and nonwords elicit higher amplitudes than high-frequency words. In both studies, the stimuli could be either words containing a bound root (e.g. *submit*, *receive*), nonwords with the same prefixes and roots (*promit*, *exceive*), control morphologically complex words (e.g. *bookmark*, *muffler*) or control nonwords with no apparent morphemes (e.g. *moobkark*, *flermuf*). The rationale of the study is explained by the authors of the first study:

If readers treat the [experimental non-words (e.g. *promit*, *exceive*)] as unanalyzed wholes, then the non-words should elicit larger amplitude N400s than the real words. This follows from the fact that whereas the words are relatively frequent in English, the bound-stem non-words (taken as a whole unit) are non-words with zero frequency. Conversely, if readers decompose the words and non-words into their constituent morphemes, and if these morphemes are represented in the mental lexicon, then the words and non-words might elicit similar amplitude N400s; both the words and non-words are made up of morphemes with the same frequency.

McKinnon *et al.* (2003)

In that first study, the results show N400 amplitudes that are significantly larger than those observed for words only for control nonwords, which means that all morphologically structured stimuli have similar N400 responses, even nonwords with bound roots. However, the second study conducted by Coch *et al.* (2013), using a more controlled methodology and more participants (16, while McKinnon *et al.* (2003) had 12), do find differences in N400 responses between items with free roots and those with bound bases. They interpret these results as an indication that morpho-orthographic decomposition might occur before 400ms, and that it is probably the later process of concatenation of the different component products of that decomposition that occurs at 400ms.⁶⁶ Therefore, the evidence for ERP is inconclusive at this stage and further replications of those studies would be welcome.

3.3.1.4. *Semantic relics*

As I have mentioned previously, all “opaque” prefixed words are not opaque to the same degree. Certain roots and prefixes do have some identifiable meaning, even if it is sometimes quite abstract. For example, the roots found in the words in (43) have a vague meaning which can be found in the different words which contain them (J.-M. Fournier 1996).

⁶⁶ However, their behavioural results are consistent with the literature on lexical decision tasks, as they find that items with bound or free morphemes differ from control items but not necessarily from each other.

- (43) a. spect ‘sight, looking’: aspect, circumspect, expect, inspect, introspect, prospect, respect, retrospect, suspect (+ spectacle)
- b. rupt ‘breaking’: abrupt, corrupt, disrupt, erupt, interrupt (+ bankrupt, rupture)

Likewise, certain productive prefixes can be found attached to bound roots with their usual meaning, such as those in (44a), while others have meanings which are recurrent but are not those found in the productive use of homographic prefixes (44b, c).⁶⁷

- (44) a. acephalous, decrease, include, intersperse, reflate, submerge
- b. rebound ‘to bound or spring **back** from force of impact’
 recall ‘to bring **back** from memory’
 reflect ‘to cast **back** (light, heat, sound, etc.) from a surface’
 repel ‘to drive or force **back** (an assailant, invader, etc.)’
 retract ‘to draw **back** or in’
- c. exclaim ‘to cry **out** or speak suddenly and vehemently’
 exhume ‘to dig **out** of the earth’
 expand ‘to spread or stretch **out**; unfold’
 expel ‘to drive or force **out** or away; discharge or eject’
 extend ‘to stretch **out**; draw **out** to the full length’

In a more formal framework, Baeskow (2006) proposes an analysis of the “lexical entailments” of different bound roots found in opaque prefixed words using natural speech from online corpora. She follows Dowty (1991) in assuming that thematic roles are best seen as “cluster concepts rather than discrete categories”, so that traditional thematic roles such as patient, agent or instrument are replaced with two proto-roles, defined over lexical entailments (e.g. sentience, change of state, movement). Some entailments are associated with the “proto-agent” and others with the “proto-patient”, and the number of entailments for each proto-role determines whether an argument is realized as a subject or a direct object. She proposes that the different roots that she analyses have the “skeletal thematic information” (Lieber 2004) shown in (45), in which <E> refers to a referential argument and the subscripts refer to the lexical entailments.

- (45) a. *duce* <E<p-agent [causation], p-patient [causally affected]>>
 b. *scribe* <E<p-agent [impulse, control, sentience, affection], p-patient [affected]>>
 c. *ceive* <E<p-agent [sentience], p-patient [affected]>>
 d. *sume* <E<p-agent [sentience, control], p-patient [affected]>>
 e. *mit* <E<p-agent [affection], p-patient [change of state, affected]>>
 f. *port* <E<p-agent [transaction], p-patient [affected, change of state]>>

Therefore, even for roots whose meaning seems the most opaque, there might be “semantic relics” that language users may be able to pick up on.

In the last sections, we have seen that opaque prefixed words display several phonological behaviours that distinguish them from simplex words, that there is some

⁶⁷ The definitions are taken from Dictionary.com (URL: <https://www.dictionary.com/>).

morphological evidence that the roots that are found in different words seem to be the same unit, that the available psycholinguistic evidence shows that they pattern with words with productive prefixes in experimental designs that neutralize semantics, and that the formatives found in such words may actually not be entirely opaque. This suggests that it is worth investigating the phonological behaviour of those words and that semantically underspecified formatives may be relevant to the phonology.

3.3.2. *Complex proper names*

While I was reviewing the literature on opaque prefixed words, I came across a paper written by Köhnlein (2015) on Dutch place names. In this paper, he argues that toponyms such as *Amsterdam* or *Wageningen* show phonological properties that are typical of morphologically complex words (e.g. extrafenestral stress, phonotactics), despite the semantic underspecification of the constituents involved in the construction of such names. It seemed to me that the same kind of properties could probably be found in English place and family names such as *Washington* or *Cambridge*, which contain recurrent elements. In 2022, I started looking into the phonological properties of those words more closely, and what started out as a little ‘side-project’ ended up being a large study – quite likely the most extensive to date – on the issue. Indeed, research on the phonology of proper names is scarce, and this is also true of the phonological properties of English proper names, on which I had difficulties finding published research. I will summarize some of the main points of that study, which is now published in *Anglophonia* (Dabouis 2023d).

The aim of the study was to see if complex names display phonological characteristics that differ from those of morphologically simplex words and resemble those of certain types of morphologically complex words which J. Kaye (1995) qualifies as having “analytic morphology”. In English, those are essentially compounds and words containing neutral affixes. I have identified eight characteristics relating to stress, phonotactics or vocalic behaviour that may tease apart simplex words from complex words. Those characteristics are shown in Table 11.

		Simplex	Neutral suffix	Compound
Primary stress	> 2 syllables, closed penult	Penult	Possibly before penult	
	> 3 syllables	Within final three-syllable window	Possibly extrafenestral	
Phonotactics	Medial clusters	Restricted	Unrestricted	
	Preconsonantal stressless /ɪj/	No	Possible	Yes
	/ŋ/	— {C[+velar], #}	Unrestricted	
	Internal superheavy syllable	Strongly restricted	Possible	
Trisyllabic Shortening		Yes	No (if the base has a long vowel)	
Vowel reduction in final constituent		Yes		No

 Table 11. The phonological characteristics investigated in Dabouis (2023d)⁶⁸

In order to study those characteristics, I manually collected 3,579 complex proper names, defined as having at least one recurrent element (most of them being final constituents well-identified by onomasticians; e.g. Watts 2004), in Wells (2008).⁶⁹ The different characteristics detailed in Table 11 were then coded, along with factors that may interact with them such as the recurrence of constituents within the dataset or the free or bound character of the constituents. The latter was established quite loosely due to the semantic opacity that characterizes those words. For example, *Appleby* (a family name or a toponym) can be analysed as *Apple* + *-by* (cp. *Allenby*, *Hornby*, *Willoughby*), and there is a freestanding word *apple* that is homographic with the first constituent. Therefore, that first constituent was treated as ‘free’, even though the name *Appleby* is semantically unrelated to the word *apple*.

Let us now review the main results of that study for each of the eight characteristics under scrutiny⁷⁰:

➤ **Primary stress:**

- Words longer than two syllables with a closed penult: primary stress is on the antepenult in the vast majority of cases (93%), regardless of the nature of the coda of the penult (e.g. *Báskerville*, *Édmondson*, *Hámington*, *Níckleby*, *Wáterstone*). If anything, having a closed penult seems to favour antepenultimate stress, as words that have an open penult have that stress pattern in only 84% of cases. Thus, those names pattern more with suffixed words or compounds (e.g. *développement*,

⁶⁸ See references cited in Dabouis (2023d) on what those characteristics are exactly and how simplex and complex words differ.

⁶⁹ The full dataset is available on OSF: <https://osf.io/t3pwy/>.

⁷⁰ This is a very rough summary. The reader is invited to consult the full paper for a more detailed and nuanced account of the results.

párenthood, búlletproof, páperwork) than with simplex words (e.g. *agénda, rèferéndum, seméster*).⁷¹

- Words that are longer than three syllables: 45 out of the 70 words in the dataset which can be pronounced with at least four syllables may be stressed on the preantepenultimate syllable (e.g. *Áddlebrough, Cánterbury, Héterington, Ándersonstown*). This is once again closer to what can be observed in complex words (e.g. *advénturousness, cháracterless, ánybody, ínfantryman*) than in simplex words, which more strongly maintain primary stress in a final three-syllable window.
- **Phonotactics:**
 - Word-medial clusters: more than a third of the dataset contains a cluster that does not occur in simplex words (e.g. *Bloomsbury /mzb/, Northbrook /θb/, Shaftesbury /ftsb/*) but may occur in complex words (e.g. /θb/ is found in *clothbound, deathbed, toothbrush*). Such clusters are more common in words whose constituents are bound than those in which they are free.
 - Preconsonantal unstressed /ɪj/: 38 out of 128 words which contain a grapheme which can represent /ɪj/ or /ɪ/ at the end of the first constituent have /ɪj/ (e.g. *Dix/ɪj/land, Holl/ɪj/wood*), while the remaining ones have /ɪ/ or /ə/ (e.g. *Harr/ɪ/son, Mar/ɪ/land*). It was found that the presence of /ɪj/ strongly correlates with the free character of the first constituent and the presence of a full vowel in the second constituent.
 - The distribution of /ŋ/: 141 words contain an illegal cluster involving /ŋ/ followed by a consonant that is not a velar stop (e.g. *Bassingbourn /ŋb/, Abingdon /ŋd/, Langford /ŋf/*) and 33 – all ending in *-ham* – have prevocalic occurrences of /ŋ/. Such sequences can be found in complex words (e.g. *kingly /ŋl/, lungful /ŋf/, longish /ŋɪ/, hangover /ŋəw/*) but not in simplex words.
 - Internal superheavy syllables: 753 words in the dataset contain an internal superheavy syllable (i.e. a syllable that contains either a long vowel followed by a coda or a short vowel followed by a branching coda). Some of those have rime structures that are attested in simplex words (e.g. *Swainson /'swɛjnsən/, Doulton /'dɔʊltən/* cp. *danger /'dɛjndʒə/, shoulder /'ʃɔʊldə/*; see J. Harris 1994: 69) but 69% of those superheavy syllables have structures that are unattested in simplex words word-medially (e.g. *Helmsdale /'hɛlmz.dɛjl/, Mountford /'mawnt.fəd/*). Once again, the presence of such syllables correlates with the free or bound character of the two constituents.
- **Trisyllabic Shortening**: 176 out of 252 words in which the primary stressed vowel is in the proper environment for Trisyllabic Shortening have short vowels (e.g. */a/bingdon, B/ɛ/veridge*) while 65 have diphthongs (e.g. *D/ɛj/vidson, /ɪj/denbridge*). Thus, the proportion of diphthongs is clearly higher than that reported by Deschamps (1994) for the rest of the lexicon, 8%. There too, there is an effect of the free character of the first

⁷¹ We can add a nuance here to consolidate the observation. As will be seen in Ch. 5, simplex words behave differently depending on the sublexicon that they belong to. Words that are foreign or learned follow the Latin Stress Rule as regards closed penults near systematically while this is only true of about a third of the remainder of the lexicon (what will be called §CORE). Here the rate of penultimate primary stress found in proper names is even lower (4%), so that the difference with simplex words is smaller but still significant.

constituent, as the proportion of long vowels is higher in words whose first constituent is free (41%) than those for which it is bound (19%). In the latter case, the proportion of diphthongs remains twice as high as that found in the rest of the lexicon.

- **Vowel reduction in the final constituent:** Two factors were found to be positively correlated with the presence of a full vowel in the final constituent:
 - a low recurrence of that constituent (e.g. *-hall* is found in 13 names and is almost systematically realized /-ho:l/ while *-ton* occurs in over 400 words and is almost systematically realized /-tən/);
 - the presence of a reduction-blocking context (e.g. a digraph, <VCe> at the end of the constituent; see Deschamps 1994: 219).

Overall, more than half of the data has characteristics that pattern with those of complex words. This could be taken as a cue that such words are often analysed as complex words. However, because there is variation among simplex words for some of those characteristics, it is possible that having one of them does not necessarily entail that language users tend to treat those words as complex. Conversely, having none of those properties may not necessarily imply that a word will not be treated as complex, as the phonology of complex words does not systematically diverge from that of simplex words.

Let us conclude with two of the implications of those results. First, it was observed for several properties that certain characteristics were more likely to be found if the first constituent was free. This can be taken to suggest that a relationship with freestanding words is made, even in the absence of a semantic connection (e.g. between *open* and the first constituent of *Openshaw*, a suburb in Manchester). This is reminiscent of what Jackendoff (2010) calls “**strawberry morphemes**”, i.e. “real words within compounds that play no role in the compound’s meaning” (e.g. *strawberry*, *cottage cheese*, *horseradish*, *sidekick*, *airplane*). While some bound constituents may be analysed as “**cranberry morphemes**”, i.e. unique bound elements with no identifiable meaning (e.g. *cranberry*, *dormouse*, *mulberry*, *hinterland*), others probably fall into a third category: recurrent bound elements with no identifiable meaning (e.g. *somer-* has no freestanding counterpart, and yet it appears in *Somerfield*, *Somerleyton*, *Somerset*, *Somerton* and *Somerville*).⁷² Second, the relationship between recurrence and the presence of a reduced vowel in the second constituent is consistent with what is known of the effects of frequency on reduction (Bell *et al.* 2009; Clopper & Turnbull 2018; Fidelholtz 1975).

Many aspects of the phonology of complex English proper names remain to be explored in future research and, as is often the case in scientific research, the study concludes with more unanswered questions than when it started. I have started exchanging with Björn Köhnlein and Sabine Arndt-Lappe on the idea of a joint research project on the phonology of complex names in Germanic languages. If this collaboration indeed takes place, we may get to answer some of those questions.

⁷² Unfortunately, no existing berry name seems available to illustrate that class.

3.3.3. *Learning to detect complexity in semantically opaque words*

In the first years during which we started confronting the Guierrian approach to other frameworks, we were told several times that opaque prefixed words could not play a role in the phonology of English because the absence of clear semantics makes them impossible to learn. While no arguments were presented to us to support that view, we took it upon ourselves to look at the evidence regarding the learnability of such constituents, especially with Jean-Michel Fournier, who had already worked on this issue (J.-M. Fournier 1996). One obvious argument supporting the idea that language users do detect semantically opaque structures lies in the distinct phonological behaviours that we have reviewed in the previous sections, and the psycholinguistic evidence showing that opaque prefixed words often pattern with words with free bases and more transparent semantics. Another one might be that what we have called “opaque” prefixed words are not systematically entirely opaque. As we have seen in §3.3.1.4, opaque prefixed words may have non-compositional semantics, but their prefixes and roots may be associated with recurrent semantics.

In Dabouis & J.-M. Fournier (2024), we reviewed the literature on the **acquisition of derivational morphology**. This literature shows that children start by acquiring suffixes with consistent form and meaning, starting with the most useful ones. They acquire prefixes later, and suffixes which are associated with changes in the form of their base (e.g. *-ion*, *-ity*, *-ic*) are acquired even later, after children enter elementary school. The ‘acquisition’ of a derivational affix is diagnosed through the use of the affix in novel forms, and this cannot be done with unproductive morphology. However, because of their semantic opacity and their sometimes-variable form, it is quite likely that the ability to identify morphological complexity in semantically opaque words emerges quite late, if at all. There are elements pointing in that direction from psycholinguistics, as recent studies on morpho-orthographic segmentation show that semantically-blind segmentation of all (apparently) morphologically structured stimuli in the early stages of visual word recognition emerges late (Rastle 2019; Marelli *et al.* 2020; Beyersmann & Grainger 2023), probably during adolescence (Dawson *et al.* 2018). Given their opacity, it would not be surprising if some language users never detected such structures, and we can further posit that the more opaque the semantics and the more variable the form, the later opaque formatives are detected, and the fewer language users detect them.

If we go beyond the evidence reviewed in previous sections, there is one other key factor that may be invoked for the identification of semantically opaque formatives: **distributional recurrence**. This has been suggested in the literature for prefixed words and complex Dutch place names (Köhnlein 2015). Naturally, we were aware of J.-M. Fournier’s (1996) proposal regarding the mechanism of “commutation” (i.e. the mechanism inherited from European structuralism which allows for the identification of linguistic units when applied to minimal pairs, morphemes and phonemes), but similar proposals have been made in the literature. Indeed, Forster & Azuma (2000) and Taft (1994) suggest that the commonality of forms like *submit*, *admit*, *permit*, *prevent*, *advent*, *circumvent*, etc. would be sufficient to create sub-units like *mit* and *vent*. Interestingly, and as was already pointed out by J.-M. Fournier, this means that even a prefix that have been non-productive for a very long time such as *se-* (which comes from Classical Latin *sē-*, where it was already archaic; Gaffiot 1934: 1409) can be detected on

the basis of its recurrence and the fact that it combines with roots which themselves combine with other prefixes, as shown in (46).

(46)	secede	accede, concede, precede, recede
	seclude	conclude, exclude, include, occlude, preclude, reclude
	secrete	accrete, concrete, discrete, excrete
	secure	procure, obscure
	seduce	adduce, conduce, deduce, educe, induce, produce, reduce, traduce
	segregate	aggregate, congregate
	select	collect, elect, intellect, neglect, recollect

The psycholinguistic literature also brings interesting intricacies regarding how this process could function for opaque prefixed words, as studies suggest a role of the recurrence of the root⁷³ (Forster & Azuma 2000; Pastizzo & Feldman 2004), but also of the cumulative frequency of the root (Taft 1979). Likewise, Wurm (1997, 2000) reports that “prefix likelihood” (i.e. “the proportion of tokens beginning with a given letter string that are prefixed”, a measure introduced by Laudanna *et al.* (1994)) and frequency could impact the processing of words. Another idea was suggested to us by an anonymous reviewer: it is possible that such structures are detected differently depending on syntactic categories. Guierre (1979) reports that the proportion of prefixed words among disyllables (separable prefixed words aside) varies greatly from one syntactic category to the other, the highest being verbs (77%) and the lowest being adjectives (6%). Therefore, it is possible that language users are more likely to detect prefixation in verbs than in adjectives, as this structure is more common in the former category than in the latter.

3.3.4. *The lexical and grammatical knowledge about non-compositional words*

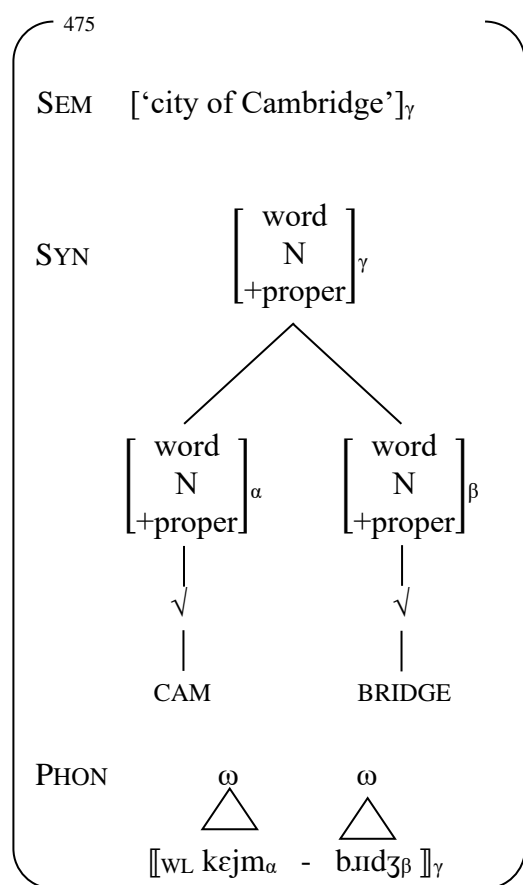
As was mentioned above, most models of phonology rely on a morpheme-based view of morphology, in which morphemes are minimal signs. This is possibly the reason why the phonological specificities of words containing more opaque formatives have not been studied as extensively as those with more productive morphology. In the two papers on which the previous sections mainly relied, when it came to formalizing the analysis, we took over formalism from Stratal Phonology, which relies on a morpheme-based approach to morphology. As there is probably a considerable amount of variability and gradience in how different language users perceive the complexity of opaque complex words and the extent to which it plays a role in their morphophonology, the analysis which we proposed in those papers – which I will discuss in the rest of this section – is best seen as a reductionist approach that seeks to identify the characteristics of some of the possible configurations.

In the previous section, we have put forward a few hypotheses regarding how language users develop the ability to detect opaque morphological structures, but a question that one should ask is what exactly it is that they learn. We argue that what is learned are lexical entries which involve complexity at certain levels, but not all, along with generalizations that capture the

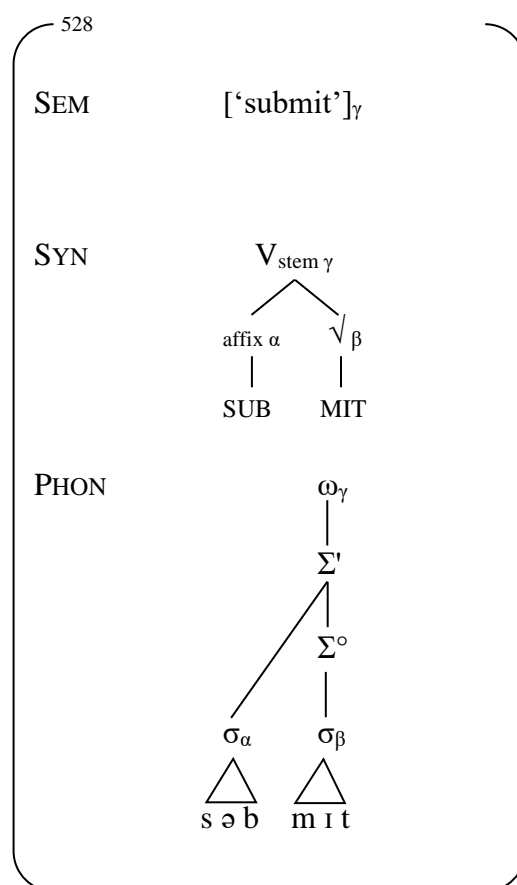
⁷³ This might be seen as a morphological family size effect.

regularities that connect those different levels of representation. At the phonological level, we saw that complex proper names often have behaviours that resemble those of words with ‘neutral’ suffixes or compounds. In Stratal Phonology, those types of units undergo word-level phonology and they are stored analytically, if stored at all. In that model, the pieces of phonological information that are stored may or may not be related to other levels of representation, morphosyntactic or semantic. This is called “coindexation” (Jackendoff 1997, 2002). Let us illustrate how this could work with examples in which I will assume that the morphological constituents have been identified but no meaning has been associated with them, shown in (47).

(47) a. Lexical entry for *Cambridge*



b. Lexical entry for *submit*

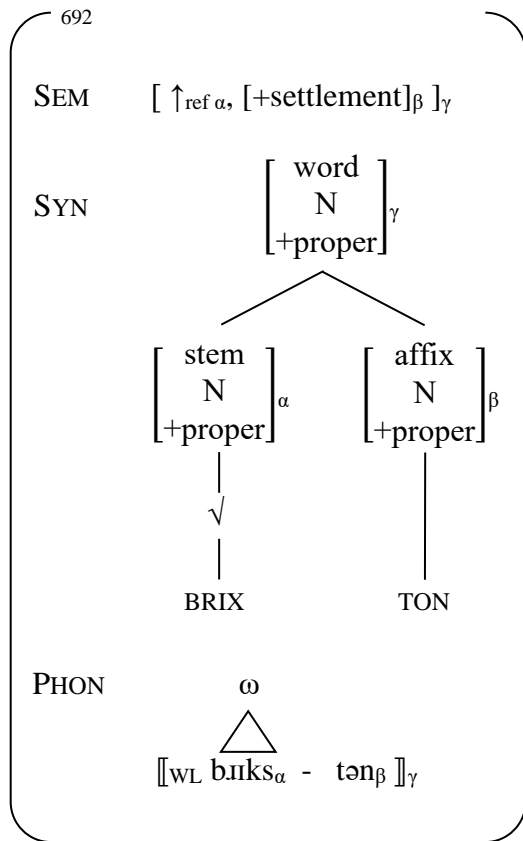


In (47), coindexation is shown using postsubscripted Greek letters (α, β, γ). For example, in (47a), the phonological string (kejm)_ω is coindexed with the morphosyntactic constituent CAM but is not coindexed to pieces of semantic representation that are smaller than that of the whole entry (‘city of Cambridge’). In (47b), the phonological component of the lexical entry is fully prosodified as this is a stem-level unit, stored nonanalytically. In this case there can still be coindexation between the two syllables and the formatives SUB and MIT but, once again, those elements are not coindexed with subparts of the semantic representation. This is the option adopted by Mascaró (2016) for Central Catalan place-names.

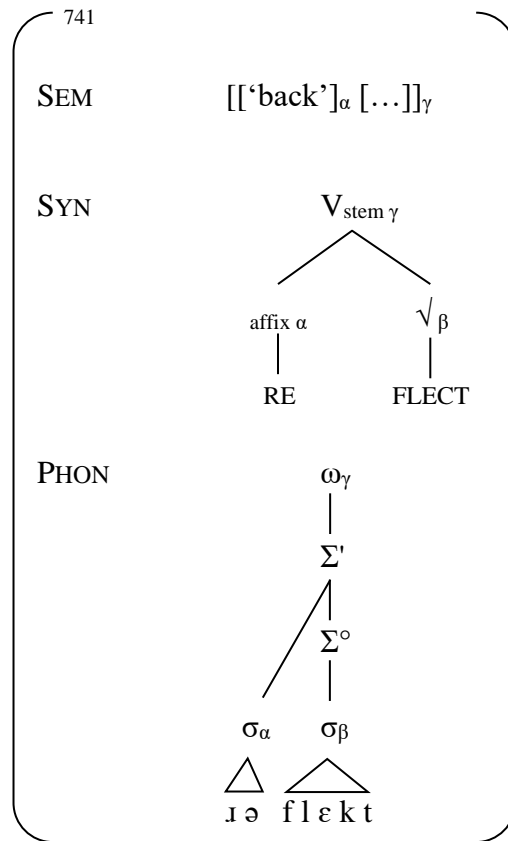
As was seen above (see §3.3.1.4), the semantic contribution of the formatives can be rather underspecified and vague, but it can be assumed that, when a language user knows a sufficient number of words sharing a formative, they may identify that meaning, leading to a

restructuring of lexical entries so that coindexation extends to the semantic level, so that some elements of the meaning of the whole word are related to some elements of the morphosyntactic and phonological representation. This can be seen in (48), where some or all the semantic representation is coindexed with the other levels.

(48) a. Lexical entry for *Brixton*



b. Lexical entry for *reflect*



In (48a), we follow the analysis proposed by Köhnlein (2015) on Dutch proper names, in which the final constituent functions as a classifier marking that the word refers to a settlement, while the first constituent is assumed to have a ‘referential pointer’ to a unique object in the world (represented as ↑_{ref}) as its only semantic content. In (48b), we represent the semantic contribution of *re-* in *reflect* through coindexation (noted α), a nonanalytically listed entry with a fully prosodified phonological representation, but the rest of the semantic representation (simply represented as “...”) is not coindexed with any formative. Note that a similar analysis could be used for words involving “strawberry” or “cranberry morphemes” (e.g. *cottage cheese*, *mulberry*), so that only one of the formatives is related to some semantic representation (e.g. CHEESE ↔ ‘cheese’, BERRY ↔ ‘berry’) but that the rest of the semantics of the word can only be coindexed with the full word.

Finally, the generalizations that are learned along with those kinds of lexical entries may or may not be specific to semantically underspecified complex words. In the case of proper names, I have argued that such words basically follow the same generalizations as words with ‘neutral’ suffixes and compounds. However, in the case of prefixed words, the phonological behaviours that are observed for rather opaque cases may be different from those of more

transparent words. There is an abundant literature on the phonological behaviours of transparent prefixed words, and this is also something that I have worked on, as mentioned at the end of §3.2.1. A very condensed and simplificatory summary of the differences that can be found between the two poles of the transparency gradient in prefixed words is shown in Table 12.

Type of prefixed word		Opaque	Transparent
Phenomenon			
Primary stress	in verbs	On base in verbs without stress-affecting suffixes (may generate patterns different from those found in simplex words)	On base (may generate patterns different from those found in simplex words)
	in nouns	Trisyllables in <i>-ate</i> stressed /102/ (like non-prefixed verbs)	Trisyllables in <i>-ate</i> stressed /201/ (unlike non-prefixed verbs)
Secondary stress in the initial pretonic position		Extremely rare	Very common
Monosyllabic prefixes ending with a vowel		Short in shortening contexts	Long regardless of context
Aspiration of base-initial voiceless stops after a prefix-final /s/		No	Yes
Gemination at the prefix-base boundary		Unattested	Possible

Table 12. The phonological characteristics of opaque and transparent prefixed words

Those observations suggest that, for some characteristics, more opaque prefixed words pattern with simplex words and do not require a special treatment and, for others, they do. Propositions for the treatment of primary stress in verbs will be made in §3.4.2.1.

Much remains to be done to improve our understanding of how opaque morphological structures function in the linguistic system of English language users. More evidence from psycholinguistics would be welcome to corroborate existing findings, and it would be particularly important to control for the diversity of semantic configurations found in prefixed words. It would also be interesting to know how complex proper names are processed. Computational approaches such as analogical models (Arndt-Lappe 2015; Skousen 2002) or Naïve Discriminative Learning (Baayen et al. 2011) may also inform us on the status of prefixes and roots as cues for the computation of stress. We have started exploring those tools in the ERSaF project. Distributional semantics and semantic vectors (Widdows & Cohen 2010) may also inform us on the intricate patterns of semantic relatedness that can be found among words sharing an apparently opaque constituent, and this may lead us to rethink transparency and compositionality in more gradient terms.

3.4. Stress and syntactic categories

In previous sections, we saw that certain factors that may affect the position of stress are sensitive to syntactic categories. Those factors will be summarized and repeated in this section for clarity. In §3.4.1, I lay out the existing claims regarding how words belonging to different syntactic categories are stressed differently. The next two sections focus on stress in verbs (§3.4.2) and nouns (§3.4.3), the most recent findings coming from the ERSaF project. Then, I will discuss the main findings of a judgement task conducted for the master's thesis of a student that I supervised, Youssef Fadel, and which was replicated for the ERSaF project (§3.4.4). Finally, I discuss some of the work I have taken part to regarding stress in disyllabic noun-verb pairs (§3.4.5). The focus will be on empirical findings. Category-specific vocalic behaviours are discussed in §4.5.3.

3.4.1. Category-specific stress?

As was seen in §2.1.3, most analyses of English stress in simplex nouns and verbs since *SPE* assume that words that belong to different syntactic categories obey different stress rules, the strongest difference being between nouns and verbs. Those are the two classes on which I will focus here. How exactly that difference manifests is more controversial, as syntactic category has been claimed to interact with factors such as quantity-sensitivity or morphology. Let us briefly review those claims.

The most common analysis, which is mostly taken over from *SPE*, is that verbs and nouns are affected by syllabic structure in different ways (see Alber (2020) for a recent overview). Both categories are affected by the presence of heavy syllables near their right edge, but the relevant syllable is the final syllable for verbs and the penult for nouns.⁷⁴ In both cases, if that syllable is heavy, it should receive primary stress, and if it is not, the preceding syllable receives primary stress. In subsequent work, there are different ways that this has been formalized. For example:

- weight-sensitive foot construction from the right edge, with extrametrical material: the final consonant in verbs, the final syllable in nouns (Hayes 1982);
- weight-sensitive foot construction from the right edge, with the assumption that verbs have a final catalectic syllable (i.e. with an empty nucleus), so that nouns and verbs are both subjected to final syllable extrametricality (Hammond 1999);
- different cophologies for verbs and nouns within OT, that differ on the ranking of the constraint penalizing footed final syllables and the one requiring syllables to be footed (Bermúdez-Otero & McMahon 2006).

In *SPE*, there was also the proposal that certain “complex verbs”, which roughly correspond to the opaque prefixed words that have been discussed in §3.3.1, are stressed differently from simplex verbs. The implicit assumption here is that nouns do not have the same behaviour. This generalization was taken over in the Guierrian School, and prefixed nouns are

⁷⁴ As was seen in §3.3.1.1, long verbs, this applies differently in long verbs, which usually have antepenultimate primary stress if the final syllable is heavy, unless the final syllable is the root in a prefixed word.

indeed found to pattern differently from verbs, adjectives and adverbs (see §3.3.1.1), which fall under what J.-M. Fournier (2007) calls the “Germanic Law”.

Another proposal that we encounter less often is that “verbs prefer final stress” while “nouns prefer penultimate or antepenultimate stress” (Moore-Cantwell 2020). In Dabouis & J.-M. Fournier (2023), we discuss this point and how weak the evidence that is cited to support that assertion is. There is some experimental evidence (see Table 13 below) pointing to a probabilistic difference between nouns and verbs, although not all the potential determinants of stress are controlled in each study.

In the course of the ERSaF project, we have reviewed as many studies as possible studying stress in verbs and nouns, either using real words or pseudo-words and using different types of methodologies (corpus studies, judgement tasks, elicitation tasks). All the different predictors discussed earlier in this section have been tested in various studies, sometimes alongside one another. The findings of those studies are synthesized in Table 13.

Study	Type	Number of syllables	Effects found				
			Rime weight	Onset weight	Prefixation	Syntactic category	Other
Kelly & Bock (1988)	Elicitation	2				✓	
Rastle & Coltheart (2000)	Elicitation	2			✓		
Guion <i>et al.</i> (2003)	Elicitation, Judgement	2	✓			✓	Analogy
Kelly (2004)	Corpus	2		✓	✓		
	Elicitation	2		✓	~		
Ryan (2014)	Corpus	2	✓	✓			
Turcsan & Herment (2015)	Elicitation	2	✓		✓	✓	Place of articulation of coda
Ktori <i>et al.</i> (2016)	Elicitation	2			✓		
Ktori <i>et al.</i> (2018)	Elicitation	2	✓		✓	✓	Orthographic weight
Treiman <i>et al.</i> (2020)	Elicitation, Judgement, Corpus	2	✓	✓	✓	✓	
Baker & Smith (1976)	Elicitation	2+	✓			✓	
Guierre (1979)	Corpus	2+	✓		✓	✓	
Ernestus & Neijt (2008)	Judgement, Corpus	3-4					Length
Domahs <i>et al.</i> (2014)	Elicitation, Corpus	3	✓				
Moore-Cantwell (2020)	Corpus, Elicitation	3+	✓			×	Final nucleus
Dabouis & J.-M. Fournier (2023)	Corpus (verbs)	2+	✓		✓		

Table 13. Empirical studies of noun and verb stress in English

What Table 13 shows is that effects of those four factors are usually found when tested, but that there is a scarcity of studies that have tested all four together, and so it is possible that the results reported by some of those studies are biased by the absence of inclusion of those possible confounding factors. Teasing those factors apart is one of the goals of the ERSaF project.

3.4.2. Primary stress in verbs

3.4.2.1. A comprehensive dictionary study

When I joined the LLL in 2011 as a master's student, the morphophonology team was working on stress in prefixed disyllabic verb-noun pairs (see §3.4.5). We eventually focused on verbs alone along with Jean-Michel Fournier, helped by colleagues at different steps in the process, notably Isabelle Girard, Pierre Fournier, Nicolas Trapateau, Marjolaine Martin and Nicola Lampitelli. We presented our work at different conferences between 2015 and 2017, and kept refining the study after receiving the input of the audiences of those conferences up until 2023, when the study appeared as a chapter in the collective book *New Perspectives on English Word Stress* (Dabouis & Fournier 2023). The presentation of the study reported in this section mainly relies on that chapter.

The aim of our study was twofold. First, we wanted to give a comprehensive overview of primary stress placement in English verbs. This implies disentangling the effects of various types of morphological structures and the rules that regulate stress in simplex verbs. Second, we wanted to question the often-repeated claim that the position of stress in verbs depends on the weight of the final syllable, a claim that is often illustrated using opaque prefixed verbs. *SPE*'s analysis included both a claim on syllable weight ("strong" and "weak" clusters in their framework) and a claim on prefixation, as was seen in the previous section, but the subsequent literature has often abandoned the claim on prefixation. For clarity, let us repeat the two generalizations under consideration in (49).

(49) a. **Weight-based stress assignment in verbs**

If the final syllable of a verb is heavy (\bar{V} or VC – under final consonant extrametricality), then it is stressed (e.g. *maintáin*, *tormént*, *usúrþ*). Otherwise, the penult is stressed (e.g. *astónish*, *édit*, *imáagine*).

b. **Prefix-based stress assignment in verbs**

Overlook the prefix(es) and assign stress normally to what remains.

Let us add some detail to those claims. First, the weight-based generalization in (49a) behaves differently in verbs that are longer than two syllables, so that they receive antepenultimate stress if their final syllable is heavy (e.g. *ánalyse*, *cómplicate*, *clárfy*). In *SPE*, this was phrased as the ASR (see (40)), which was blocked if an = boundary was found between the last two syllables (e.g. *comprehénd*, *intervéne*, *introduíce*).⁷⁵ As for, (49b), it corresponds to what J.-M.

⁷⁵ We find the same distinction between heavy-final verbs whose last syllable is the root and other verbs in the later literature (Hayes 1982; Liberman & Prince 1977), with the former having final stress and the latter antepenultimate stress.

Fournier (2007) calls the “Germanic Law”. Here “normally” refers to his “Normal Stress Rule”: initial stress in disyllables and antepenultimate stress in longer words. As it is possible that the two generalizations largely overlap as many opaque prefixed verbs have a final heavy syllable, we wanted to see what the evidence was for those two claims using a large dictionary dataset.

Thus, we used the LLL’s Dictionary Database to extract all 5,236 words marked as being verbs (this is based on the Macquarie Dictionary), using the British pronunciations given by Jones (2006). We excluded 293 entries that are marked as being rare, obsolete or belonging to another variety of English in the *OED* or having no entry at all in that dictionary. Another necessary step to study verbs was to take out entries that are in fact conversions, as they may be preserving the stress from their nominal base and may thus bias the results. Thus, I read literature on conversion (Bauer *et al.* 2013; Bram 2011; Lieber 2004; Plag 1999, 2003), and found out that no single criterion could be used to properly identify converted words. Therefore, based on that literature, we used four criteria to make our classification: dates of first attestation, frequency, semantic range and semantic dependency.⁷⁶ This allowed us to identify 1,395 cases of conversion, 1,312 from nouns (e.g. *author, catalogue, exile, patent, tapestry*) and 94 from other categories (e.g. *busy, even, farewell, narrow, opaque*). Thus, our final dataset contains 3,548 verbs, among which 1,918 are disyllables, and 1,630 are longer. Then, we annotated the data morphologically so as to be able to analyse different morphological categories separately. This was done manually, relying on information taken from the *OED*. We identified:

- **Prefixed words:** Prefixed verbs either with compositional semantics and no category change (e.g. *disprove, rewrite, unlock*) or more opaque ones, identified through etymology (e.g. *compile, depose, react*). Verbs that have an initial adverbial particle (e.g. *backfire, foreclose, outdate*) are treated along with prefixed words. Among opaque constructions, we distinguished words in which the base is free from those in which it is bound;
- **Suffixed words:** Suffixed verbs identified through etymology, as was done for prefixed verbs. They are further sorted based on whether they contain known stress-shifting suffixes (mainly *-ate* and *-ify*) or stress-preserving suffixes (e.g. *-en, -er, -le*). Verbs containing the suffix *-ize*, which is known to oscillate between those two behaviours (Bermúdez-Otero 2018b; Guierre 1979: 711-714; Selkirk 1982; Szpyra 1989: §2.2.3; Trevian 2015: §13.2) are treated separately;
- **Compounds:** all verbs formed from the association of two free bases (e.g. *blindfold, skyjack, cross-examine*).
- **Simplex:** No identifiable morphology (e.g. *argue, harangue, listen*).
- **Other:** Words with other types of morphological structures or more obscure cases, but that are not numerous enough to form a homogenous category (e.g. *animadvert, choreograph, genuflect, flabbergast*).

Those categories are distributed as shown in Table 14.

⁷⁶ Semantic range was measured by the number of meanings given for a member of the pair in the *OED*, and the member with the highest number of meanings was assumed to be more likely to be the base. As for semantic dependency, we checked whether the meaning of a member of the pair was dependent on that of the other and fits into one of the categories of semantic relationships found in productive conversion listed by Lieber (2004: 92) and Plag (1999: 220).

	Disyllables		Long verbs		TOTAL	
Compositional left-edge constructions	227	12%	362	22%	589	17%
Opaque prefixed constructions - free base	364	19%	110	7%	474	13%
Opaque prefixed constructions - bound base	665	35%	105	6%	770	22%
Simplex	278	14%	15	1%	293	8%
Bound base + suffix	112	6%	18	1%	130	3%
Stress-shifting suffixes	40	2%	617	38%	657	19%
Stress-preserving suffixes - free base	114	6%	6	<1%	120	4%
-ize - free base	2	<1%	334	20%	336	9%
Compounds	88	4%	24	2%	112	3%
Other	28	1%	39	2%	67	2%
TOTAL	1,918		1,630		3,548	

Table 14. Morphological distribution in the dataset used in Dabouis & J.-M. Fournier (2023)

The figures in Table 14 reveal interesting facts about the morphology of verbs. In disyllables, we find that 66% of words are ‘left-edge’ constructions (compositional constructions or opaque prefixed words), while longer verbs are suffixed in 60% of cases and ‘left-edge’ constructions in 35% of cases, with an additional 20% if we include verbs that are both prefixed and suffixed (e.g. *abdicate*, *complicate*, *objectify*, *recognize*). This confirms the observations that verbs are heavily built on the left mentioned in §3.3.3, and that this could partly explain the role that prefixation plays in stress assignment in that category.

Let us now briefly review the results. In compositional left-edge constructions, we found that primary stress is on the root in 99% of cases (e.g. *dèclássify*, *rèwrite*, *òutbáalance*). Verbs that contain stress-shifting suffixes (mainly *-ate* and *-ify/-efy*) largely comply with the stress patterns that are associated with those suffixes, which is most of the time antepenultimate primary stress (e.g. *colláborate*, *persónify*, *rárefy*). Almost all verbs that contain neutral suffixes also show the typical behaviour of that type of word: primary stress is on the same syllable as in the base (e.g. *awáke* ♦ *awáken*; *búrn* ♦ *búrnish*; *frésh* ♦ *freshen*), and the only exceptions are cases in which the semantic relationship between the base and the derivative has become quite opaque (e.g. *bómb* ♦ *bombárd*; *ímage* ♦ *imáagine*). As found in the previous literature, the verbs that contain the suffix *-ize* almost systematically have primary stress on the same syllable as their base, which may generate pre-antepenultimate patterns (e.g. *concéptualize* ♦ *concéptual*; *fédéralize* ♦ *federal*; *márginalize* ♦ *márginal*). When stress is not on the same syllable as that of the base, the semantic relationship between the base and the derivative is sometimes quite obscure (e.g. *canál* ♦ *cánalize*; *vólatile* ♦ *volátitalize*; *Galváni* ♦ *gálvanize*).

When reviewing the literature on stress in verbs, I did not find any studies of verbal compounds. This is probably not surprising considering the small number of such words. Therefore, I read the recent literature on stress in nominal compounds, which have been extensively studied (Bell & Plag 2013; Giegerich 2009; Plag 2006; Plag *et al.* 2007, 2008), so as to identify the potential determinants of stress in verbal compounds.⁷⁷ Thus, we tested factors

⁷⁷ We also had to exclude some of those data: 26 compounds that are actually backformations, and which systematically reproduce the stress pattern of their base (e.g. *ghóstwrite* ♦ *ghóstwriter*; *jáywalk* ♦ *jáywalking*).

that were found to be relevant in noun-noun compounds: lexicalization (using spelling and compound frequency), informativity (using the frequency of the two constituents and synsets⁷⁸) and word length (different measures were tested: number of syllables of each constituent, of the whole compound, or of the sequence after the main stress of the first constituent). We also included an additional factor, the syntactic category of each constituent, as we thought that it might be relevant.⁷⁹ Those factors were tested in a binary logistic regression on the 112 compounds found in our dataset, and only two were found to be significant predictors of the position of primary stress in the compound: spelling (one-word vs. hyphenated) and the syntactic category of the first constituent. The distribution of stress patterns in the data depending on those two factors is shown in Figure 15.

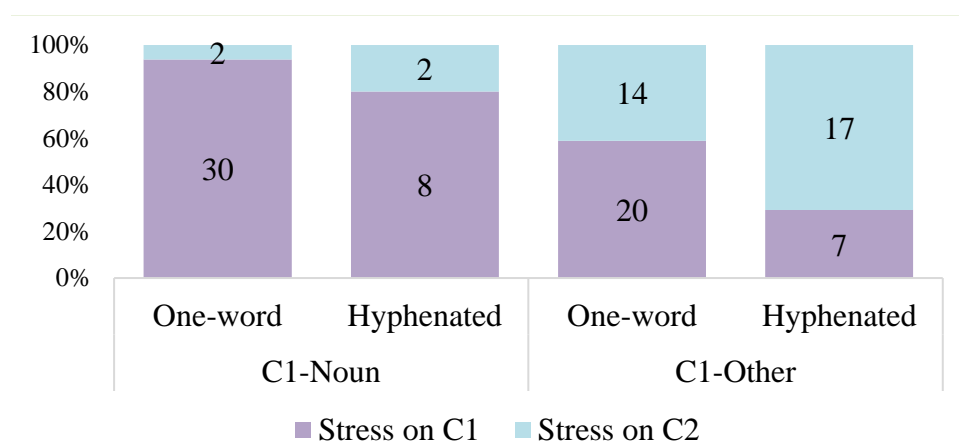


Figure 15. Position of main stress in verbal compounds depending on the syntactic category of the first constituent and spelling (taken from Dabouis & J.-M. Fournier 2023)

As can be seen in Figure 15, we find more primary stress on the first constituent if it is a noun (e.g. *fróstbite*, *pán-fry*, *wáterlog*), and if the compound is spelled as a single word (e.g. *bróadcast*, *hándcraft*, *súnburn*). Another factor that has been found to be relevant in noun-noun compounds is analogy: compounds that share a constituent tend to be stressed in the same way (see e.g. Arndt-Lappe 2011). However, our dataset is rather restricted and there are not many compounds sharing a constituent, and when they do, we do not systematically have the same stress pattern (e.g. *áir-dry*, *dríp-dry* vs. *spin-dry*). Additional studies using speech data would be welcome to complement this first study, especially considering that studies on compound stress have shown that there is a considerable amount of inter- and intra-speaker variability, which cannot be reflected in dictionary data.

Finally, we focused on the two generalizations in (49). As one motive behind our investigation of the effects of opaque prefixation on stress was to question the assumption that opaque prefixed words actually have no morphological structure, we left out the 474 words which we analysed as semantically opaque but for which the base is free, as those words may be less controversially analysed as having morphological structure. In those words, stress is on

⁷⁸ Synsets correspond to sets of words with similar meanings. A word with a greater number of synsets is assumed to be less informative and is thus expected to be less likely to be stressed.

⁷⁹ This factor was not present in the literature on nominal compounds, as existing studies focus on noun-noun compounds.

the base in 99% of cases (e.g. *accústom, become, defáult, preóccupy, redréss, ùnderstánd*). The remaining words are either opaque prefixed words with a bound base (e.g. *commit, eject, intercept, remember*), suffixed words with a bound base (e.g. *cherish, glisten, manage, sprinkle*) or simplex verbs (e.g. *argue, copy, follow, manoeuvre, orient, usurp*).

In order to test the weight-based generalization, we had to make decisions on how to measure the weight of the final syllable, as there exist several approaches. We deliberately chose to take the most favourable options for this generalization whenever different options were possible in the literature. I am leaving out the details for each choice we made, which can be found in the chapter, and will simply list here the assumptions that were adopted for weight coding, as shown in (50).

- (50)
- a. /ə/ is morales
 - b. Final unstressed /ɪj/ and /əw/ are monomoraic (but final stressed /əw/ is treated along the other long/tense vowels)
 - c. Short/lax vowels are monomoraic
 - d. Long/tense vowels are bimoraic, except /aw, oj/, which are trimoraic
 - e. The last consonant is morales
 - f. Preconsonantly, sonorants and /s/ are morales if unstressed and have one mora if stressed, to the exception of /ŋ/, which is bimoraic
 - g. Preconsonantly, other consonants are treated as monomoraic

Those assumptions led to the weights shown in Table 15.

Vowel	Final consonant(s)				
	__#	__C#	__C[+son]C# __sC#	__/ŋ/C#	__C[-son]C#
/ə/	0μ	0μ	0μ	-	1μ
unstressed /ɪj/	1μ	-	-	-	-
unstressed /əw/	1μ	-	-	-	-
/aw, oj/	3μ	3μ	3μ	-	-
/a, ε, ɪ, ʊ, ə, ə/	-	1μ	2μ ⁸⁰	3μ	2μ
/ɑj, ɛj, ɪj, ʊw, əw, aw, oː, ɑː, əː, ɪː, εː/	2μ	2μ	3μ	-	3μ

Table 15. Weight coding adopted for deradical and simplex verbs in Dabouis & J.-M. Fournier (2023). Heavy syllables are shown in grey.

Subsequently, those weights were converted into a binary measure, distinguishing light syllables (morales or monomoraic) from heavy syllables (bimoraic or trimoraic). We examined the effects of the weight of the final syllable and those of opaque prefixation in disyllables and longer verbs separately. In disyllables (n=1,055), we tested those factors in a binary logistic regression with the position of primary stress as the dependent variable. The two factors came

⁸⁰ The only attested cases have final stress.

out as highly significant predictors of the position of primary stress in those words. The distribution of the data can be seen in Figure 16.

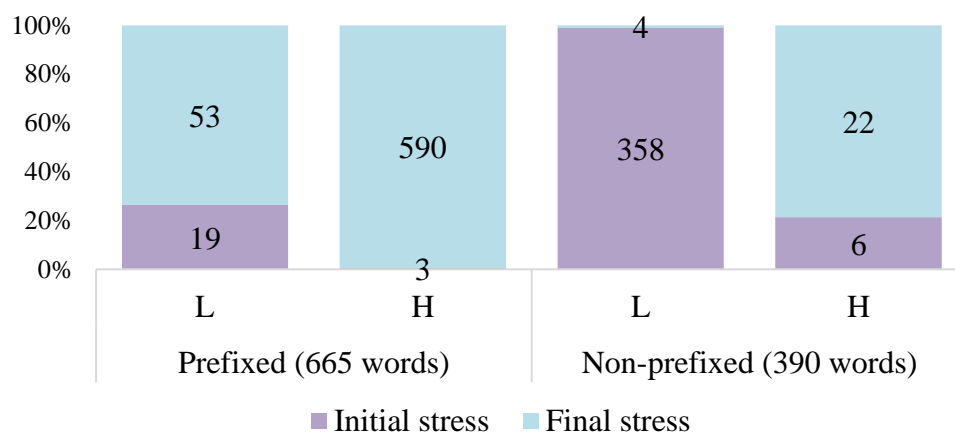


Figure 16. Position of primary stress in disyllabic deradical or simplex verbs depending on the weight of their final syllable (L = Light, H = Heavy), taken from Dabouis & J.-M. Fournier (2023)

As can be seen from Figure 16, the two central columns are those for which the distributions are the closest to categorical, and those are precisely the two configurations in which the two generalizations make the same predictions: opaque prefixed verbs with a heavy final syllable (e.g. *ejéct*, *expórt*, *retáin*) and non-prefixed verbs with a light final syllable (e.g. *cáncel*, *lísten*, *rénder*). The other two categories show the relevance of each generalization: final stress is the dominant pattern in both configurations, prefixed verbs with a final light syllable (e.g. *asséss*, *emít*, *propél*) and non-prefixed verbs with a heavy final syllable (e.g. *cajóle*, *molést*, *ordáin*). In the chapter, we note that this kind of distribution would be consistent with a probabilistic model such as MaxEnt (Goldwater & Johnson 2003), which is a version of OT which uses weighted constraints (Pater 2009, 2016) and is particularly well-suited for probabilistic processes. For example, if we assume that the grammar contains three constraints A, B and C, with weights of 3, 2 and 2, respectively, we could have the configuration shown in (51).

(51)

	A 3	B 2	C 2	<i>H</i>	p(grammar)
Candidate 1	1			-3	0.73
Candidate 2		1	1	-4	0.27

Here, Candidate 1 violates Constraint A while Candidate 2 violates Constraints B and C, yielding “harmonies” (noted *H*, i.e. the weighted sum of violations) of -3 and -4, respectively. Probabilities are then computed by calculating the proportion of the exponential of the harmony for a given candidate by the sum of the exponential of the harmonies of all candidates. Here, for Candidate 1 this yields $e(-3)/(e(-3)+e(-4)) = 0.73$.

The patterns observed in verbs could be analysed using two constraints for weight sensitivity and prefixation, which may conspire and yield greater regularity when they predict

similar outcomes. We did not develop such an analysis in the chapter, but let us sketch what it could look like. Let us posit the three constraints in (52), the first two of which are taken over from existing foot-based accounts of English stress using OT (Bermúdez-Otero & McMahon 2006; Pater 2000).

- (52) a. FTBIN
 Feet must be binary at some level of analysis (μ , σ).
- b. PARSE- σ
 All syllables must belong to feet.
- c. ALIGN-ROOT
 Align the left edge of the root with the right edge of the head of the Prosodic Word.

Let us assume that the candidates to evaluate will be footed either as ($\acute{\sigma}$ σ) or as $\sigma(\acute{\sigma})$. If we input the data shown in Figure 16 into the MaxEnt Grammar Tool (Wilson & George 2009), which is a software designed to compute the weights of the constraints of a MaxEnt grammar when provided constraint violations for the different candidates. This yields the tableau in (53).

(53)

		ALIGN-ROOT 5.06	FTBIN 2.57	PARSE- σ 1.58	H	p(grammar)	p(data)
Prefixed – L#	($\acute{\sigma}$ σ)	1			-5.06	0.29	0.26
	$\sigma(\acute{\sigma})$		1	1	-4.15	0.71	0.74
Prefixed – H#	($\acute{\sigma}$ σ)	1	1		-7.63	0.00	0.005
	$\sigma(\acute{\sigma})$			1	-1.58	1.00	0.995
Non-prefixed – L#	($\acute{\sigma}$ σ)				0.00	0.98	0.99
	$\sigma(\acute{\sigma})$		1	1	-4.15	0.02	0.01
Non-prefixed – H#	($\acute{\sigma}$ σ)		1		-2.57	0.27	0.21
	$\sigma(\acute{\sigma})$			1	-1.58	0.73	0.79

This very simplistic model can capture the data with a relatively good accuracy and shows the cumulative effects of the constraint enforcing the construction of bimoraic trochees (FTBIN) and that enforcing left-alignment of the head foot to the first syllable of the root. However, it should be refined considerably in the context of a broader analysis of stress assignment in English. For example, if we want to capture the fact that monosyllabic prefixes have reduced vowels in initial pretonic position (i.e. they are not footed), but that longer prefix sequences may bear secondary stress if there are at least two pretonic syllables (e.g. *còrrespònd*, *èntertáin*, *ìntervéne*), we may need to introduce an additional constraint that requires prefixes to be

unfooted which is dominated by a *LAPSE constraints banning sequences of two unfooted syllables. We would also need to take the weight of the first syllable into consideration, and possibly use an alignment constraint requiring the head foot to be right-aligned.

Now turning to the 138 verbs that are longer than two syllables in our dataset, we have not conducted any statistical analyses in the chapter, but tested the two generalizations and compared their respective efficiencies, which were very similar. In that subset, 20 words contain the suffix *-ize* which should be left aside on the grounds that it has been argued to be a “Long Retractor” when it attaches to roots; i.e. it imposes antepenultimate stress regardless of the weight of the penult (Bermúdez-Otero 2018b).⁸¹ Let us now revisit those results by running a statistical analysis of that same dataset using the variables that we were interested in in the chapter. Let us here focus on trisyllabic verbs and keep out the two four-syllable verbs, *reconnoitre* and *impoverish*. This leaves 116 long verbs, 100 of which are prefixed. One difficulty that we have in trisyllables is that the stress pattern resulting from the ‘stress-repelling’ effect of prefixes is different depending on how many syllables are found in the prefix sequence. Therefore, the presence of prefixes was here implemented by indicating that non-prefixed words (e.g. *bamboozle*, *gallivant*, *manifest*) have zero prefix syllables, prefixed words with a monosyllabic prefix (e.g. *deliver*, *elicit*, *replenish*) have one prefix syllable and prefixed words with either two monosyllabic prefixes or a disyllabic prefix have two prefix syllables (e.g. *apprehend*, *contradict*, *supersede*). Here, we can use conditional inferences trees, which are a kind of non-parametric tree-structure model of classification. It can be used as an alternative to logistic regression, especially in cases in which predictors interact in complex ways (Levshina 2021). This was done using the function `ctree` from the package `party` with the code shown in (54), in which `WEIGHT` refers to the weight of the final syllable as calculated following Table 15, and it returns the plot in Figure 17.

```
(54) plot(ctree(Stress ~ PrefixSyll + weight, data = data))
```

⁸¹ Although our results are not entirely consistent with that claim, as our dataset contains five root-derived words that have a heavy penultimate syllable, and they are almost equally divided among penultimate and antepenultimate stress: *amórtize*, *etérnize* vs. *éxorcize*, *fráternize*, *récognize*.

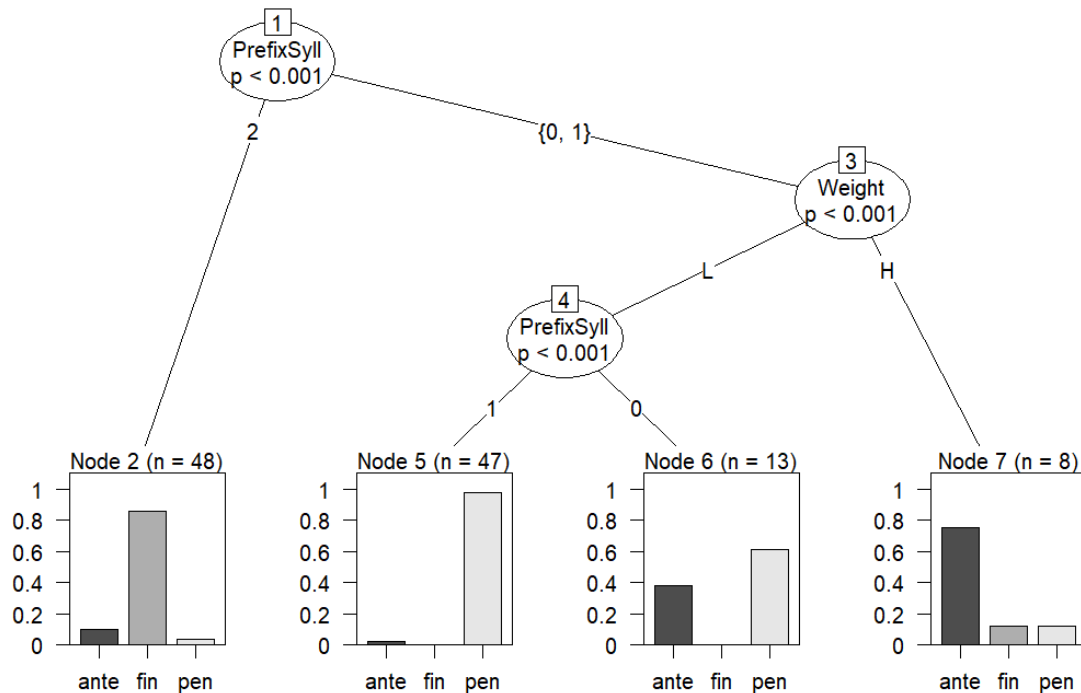


Figure 17. Conditional inference tree for primary stress in deradical and simplex trisyllabic verbs (data from Dabouis & J.-M. Fournier 2023)

In a conditional inference tree, each node represents a significant split in the data, and the branches indicate the different values for the factor in question. In Figure 17, the main split is found for the number of prefix syllables, with a distinction between words that have two such syllables (with a large predominance of final stress) and other words. The second most important split has to do with the weight of the final syllable. The eight words with a heavy final syllable show a preference for antepenultimate primary stress, regardless of whether they are prefixed (*cómpliment*, *íimplement*, *óccupy*) or not (*dámascene*, *gállivant*, *mánifest*). The two words that have another pattern are prefixed: *contínue* and *íimportúne*. The absence of a difference between prefixed and non-prefixed verbs here may have to do with the small number of relevant words. Among words with a light final syllable, penultimate stress is mainly found in words containing a monosyllabic prefix, but we do find that more than half of non-prefixed words also have that pattern (e.g. *canóoodle*, *malínger*, *sequéster*). The remaining ones have antepenultimate stress (e.g. *mánacle*, *mónitor*, *órient*). Note once again that the small number of words in that category invites caution regarding the scope of the conclusions that can be drawn from those data. Thus, in long verbs, the effects of weight are more difficult to show because of the scarcity of long simplex verbs, and we see strong effects of prefixation. We see patterns that resemble *SPE*'s ASR only if the verb does not have two prefix syllables.

As a conclusion to our chapter, we argue that both generalizations can efficiently capture the stress patterns found in our data, and that both may be required. We also underline the different theoretical ‘cost’ of each generalization: the assumptions in (50) for the weight-based generalization or the assumption that opaque prefixed words are perceived by language users as morphologically complex (see §3.3). To my knowledge, there has not been a study of verb stress that is as extensive as the one we have conducted. However, there are still aspects to be investigated in future work. First, there is a factor that I have learned about after this study was

conducted and that we did not test: the role of onsets in the computation of stress (see §2.1.3). Second, we tested the effects of weight following the commonly assumed claim that stress in verbs depends on the weight of the final syllable, but we have not looked at whether weight affects stress in other positions. Finally, as we have seen in this section, certain configurations are very hard to come by, and so other types of data may be required to study how stress should be placed in those configurations. This is what Aaron Seiler is doing for his PhD in the ERSaF project has started doing using an elicitation task using long pseudo-verbs.

3.4.2.2. *A pseudo-word study on long verbs*

In the context of the ERSaF project, Aaron Seiler conducted a production study on trisyllabic pseudo-verbs. This is a very welcome study as most pseudoword studies have focused on disyllabic words, and occasionally long nouns, as was seen in §3.4.1. My contribution to this study was mainly in offering suggestions on the stimuli used in the experiment, possible analyses of the data and classification of the different groups of stimuli. The study was presented at the 2024 OCP conference as a collective presentation of the whole project team alongside the study discussed in §3.4.4, and it will be submitted as a paper authored by Aaron Seiler and Sabine Arndt-Lappe.

In this study, the focus was on primary stress assignment in trisyllabic pseudo-verbs, and the possible predictors that were tested were syllable weight and prefixation. Syllable weight was operationalized by having the stimuli have a single heavy syllable in one of the three positions and only light syllables elsewhere. A syllable was considered heavy if it was closed, and the final consonant of the word was assumed not to contribute weight to the final syllable. Thus, a final syllable was treated as heavy only if it has two coda consonants. The main prediction from the literature (following (49a)) is that weight is only relevant in the final syllable, and that having heavy syllables in other positions should not have any effects. However, in a pseudoword production study, Domahs *et al.* (2014) found that weight effects could be observed in all positions in pseudo-nouns, which could be interpreted as a manifestation of the tendency for heavy syllables to be stressed, i.e. what is often called the WEIGHT-TO-STRESS principle. As for prefixation, stimuli could start with an orthographic sequence that is a potential prefix, some of them being productive (e.g. *pre-*, *un-*) and other non-productive (e.g. *con-*, *se-*). The prediction for that factor is that prefixes will tend not to receive primary stress (following (49b)), and that this will be more often the case for productive prefixes than non-productive ones. Table 16 shows examples of the different configurations found for the 27 stimuli used in the study.

	Non-prefixed	Prefixed	
		Non-productive	Productive
LLH	<i>caborasp</i>	<i>becarost</i>	<i>precomult</i>
LHL	<i>tufontap</i>	<i>secolpam</i>	<i>rebaltin</i>
HLL	<i>pendatip</i>	<i>contasit</i>	<i>unpavin</i>

Table 16. Examples of stimuli used in the pseudo-verbs production experiment

The participants were 50 native speakers of British English who were recruited via Prolific (<https://www.prolific.com/>). They were asked to read the verbs in carrier sentences, (e.g. *He managed to caborasp a small cabin*) after being presented with a short definition for the pseudo-verb (e.g. *If you caborasp something, you build something without anyone's help*). The responses were coded auditorily by three trained raters and a response was only kept if at least two raters agreed. Other exclusion criteria were applied to responses that were not usable (e.g. the target item was not produced in its carrier sentence, or it was produced with changes to its segmental structure), leading to the exclusion of 9.6% of responses, leaving 1081 observations.

Several statistical analyses have been conducted, and I will here focus on the conditional inference trees analysis, whose results are shown in Figure 18.

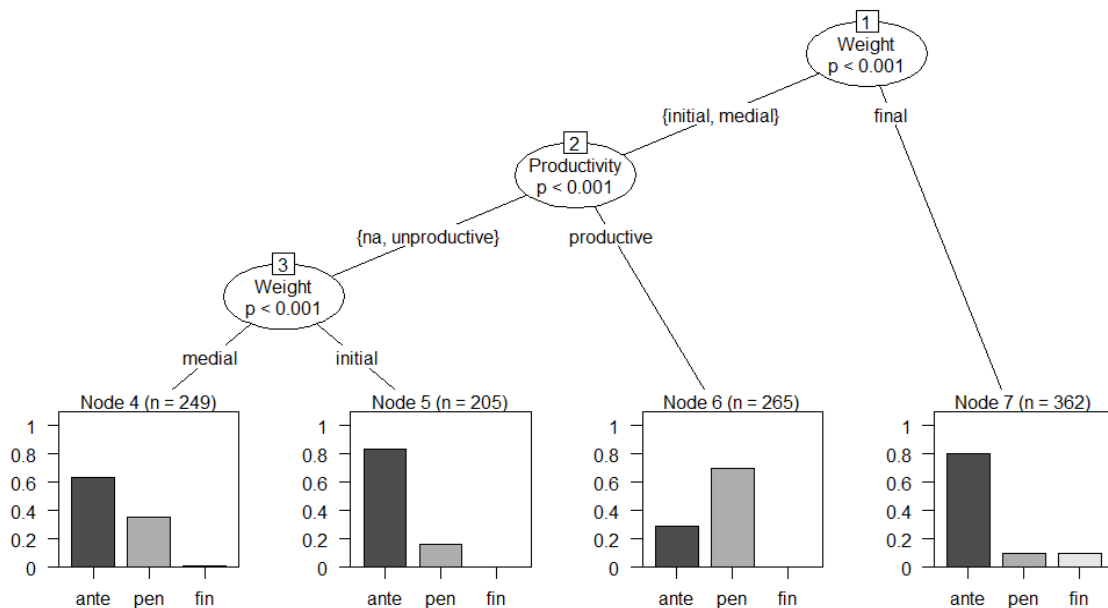


Figure 18. Conditional inference tree for primary stress in trisyllabic pseudo-verbs

The first striking result is the strong predominance of antepenultimate stress overall, as it is found in over 60% of responses. Second, syllable weight has an effect in all positions, contrary to what is usually assumed in the literature, and the stimuli with a final heavy syllable were mostly produced with antepenultimate stress. This is consistent with *SPE*'s analysis, in which such items undergo the ASR, and Domahs *et al.*'s (2014) results on pseudo-nouns. One may wonder where this high rate of antepenultimate stress comes from as there are not many verbs with no morphological structure or a final heavy syllable. It is possible that this is modelled on morphologically complex verbs, which very often have antepenultimate stress, especially among stress-shifting suffixes such as *-ate* or *-ify*. Given that pseudo-nouns with a heavy final syllable behave in the same way as our pseudo-verbs, it is also possible that the same phonological generalizations apply to both long verbs and long nouns. The results regarding prefixation are more difficult to interpret. The stress-repelling effect of productive prefixes is to be expected, but it is unexpected that no effects can be found for non-productive ones. Moreover, a closer look at the data shows that there are strong differences between prefixes that were categorized in the same way. For example, pseudo-verbs starting with *dis-* and *un-* are

almost systematically stressed on their penult while pseudo-verbs which begin with *re-*, *pre-* and *de-*, which were also categorized as productive, are more evenly distributed between penultimate and antepenultimate stress. Similar differences can be found among non-productive prefixes, with *ob-* being categorically associated with antepenultimate stress while *be-* and *con-* show sizeable proportions of penultimate stress. Thus, it appears that such a binary distinction based on productivity fails to capture the results, and more fine-grained measures of prefix-specific characteristics may be required. Those could be prefix likelihood measures (Wurm 1997, 2000), the proportion of words containing the prefix in which the base is free, or the type or token frequency of the prefix.

Those results have also been analysed using AML (“Analogical Modeling of Language”, Skousen *et al.* 2013), a computational analogical model, to test whether the stress patterns produced by participants in this experiment could be predicted, at least partly, on the basis of the dictionary data reported in the previous section (Seiler & Arndt-Lappe 2024). Their results show that using lexical support computed from the AML model as a predictor into a regression model alongside weight and prefixation improves the model. This can be taken to mean that participants did use the patterns of verbs in their mental lexicon to make their decisions on how to stress pseudo-verbs in this experiment. They also find significant effects of prefix likelihood. However, the model still underpredicts antepenultimate stress, and this can be taken to imply that participants have used words from their lexicon that are not verbs, or that their grammatical knowledge regarding verbs is somehow at odds with the lexicon.

3.4.3. Primary stress in nouns

This is a topic on which I have not worked a lot personally, but Marie Gabillet is working on this in the context of her PhD in the ERSaF project. Her work should shed light on a number of issues:

- **The role of syllable weight:** As was seen previously, the usual claim found in the literature is that the final syllable of nouns is extrametrical and that only the weight of the penult plays a role in stress assignment. However, there is some evidence that weight plays a role in different positions, including the final syllable (e.g. Domahs *et al.* 2014; Garcia & Goad in preparation). Moreover, as was seen in §2.1.3, the role of onset weight has not been investigated outside of the initial onset of disyllabic words. Finally, it is possible that words belonging to different sublexica display different sensitivities to syllabic weight.
- **The role of sublexica:** As will be seen in Ch. 5, words that belong to different sublexica (core lexicon, foreign vocabulary, learned vocabulary) display different stress patterns in dictionary data. It will be informative to establish whether that holds in experimental studies. In that chapter, a first systematic exploration of this factor using dictionary data will be reported.
- **The role of opaque prefixation:** We saw in §3.3.1.1 that nouns show sensitivity to the presence of opaque prefixes for the position of primary stress in disyllables (although less so than other syntactic categories) and for the presence of post-tonic secondary stress in the second syllable of disyllables with initial primary stress. It will be

interesting to try and replicate those findings using dictionary and experimental data and establish if such structures have an impact on stress in longer words.

- **The role of word length:** The preliminary results on verbs (§3.4.2.2) show that although disyllabic verbs with a final superheavy syllable do usually have final primary stress, this is not the case in longer verbs, which usually have antepenultimate stress. Moreover, Ernestus & Neijt (2008) report the dictionary and experimental evidence showing that trisyllables and quadrisyllables are stressed differently, with the latter being more likely to have penultimate stress than the former. Can we replicate such results regarding the effects of words length?
- **The role of final nuclei:** Moore-Cantwell (2020) reports that final /ɪj/ (which she transcribes [i]) has a greater tendency to be associated with antepenultimate primary stress than other final nuclei. In the lexicon, this is potentially correlated with sublexica, and so additional research is needed to disentangle the two. In §5.3.2, I will show that, using Moore-Cantwell's data, we were able to find that both effects appear to hold, but that the role of final nuclei is much more restricted when sublexica are taken into consideration.

Thus, there is a lot of work to do in order to better understand how all those factors affect stress, a task that is made difficult by the fact that they often overlap. For example, if we want to isolate the effects of sublexica in a production experiment, we need to hold the other predictors constant. This could involve giving different semantic cues to participants (e.g. the word is a food item from Spain, or a place name in Japan) with identical test items presented to different groups of participants. However, this is a particularly difficult task as the words belonging to different sublexica usually do not look the same: they vary in length, in phonotactics or graphotactics. Thus, the semantic information and the form of the stimuli may conflict, and it is unclear what the effects on stress might be.

3.4.4. *Disyllabic pseudo-nouns and verbs*

In 2022, as the ERSaF project was just starting, a master's student, Youssef Fadel, asked me to supervise him for his second year. He had to retake that second year and change supervisors, which meant that he had to do in one year what other students in the same master's degree normally do in two. We discussed possible topics, and we settled on stress in disyllabic verbs and nouns, to try and disentangle the different factors at play and contribute to the field as the number of studies that have controlled for many different factors is very restricted (see Table 13). The factors that we were interested in were the number of initial onset consonants, of final coda consonants, the presence of a word-initial string that may or may not be a prefix and the syntactic category of the word (noun or verb). We decided to use a judgement task, in which participants would be presented with two possible pronunciations of a pseudoword presented visually. The only study that had tested all those factors, Treiman *et al.* (2020), was a production study, and so we were interested in seeing whether the results would be different using a different paradigm. Moreover, as time was a strong restriction, it had the advantage of making the data analysis easier than with production data.

The stimuli were designed as follows. They could have zero, one or two onset consonants, one or two coda consonants, have an initial string that may or may not be a prefix and be presented in a verbal or a nominal frame. We built 4 test words per category, yielding a total of 48 test words, to which we added 40 filler words. Each word was recorded with two possible pronunciations (initial or final stress) by a native speaker of Canadian English. Only short vowels were used to keep syllable weight constant, and reduced vowels were used in unstressed syllables. The target words and their possible pronunciations are shown in Table 17.

Onset	Final rime	Overall structure	Prefixed		Non-prefixed			
Empty	VC	VCVC	<i>adet</i>	/ˈædət/	/əˈdɛt/	<i>idet</i>	/ˈɪdət/	/ɪˈdɛt/
			<i>abul</i>	/ˈæbəl/	/əˈbʌl/	<i>ibul</i>	/ˈɪbəl/	/ɪˈbʌl/
			<i>elut</i>	/ˈɛlʊt/	/ɪˈlʌt/	<i>ilut</i>	/ˈɪlʊt/	/ɪˈlʌt/
			<i>elop</i>	/ˈɛləp/	/ɪˈlɑp/	<i>ilop</i>	/ˈɪləp/	/ɪˈlɑp/
	VCC	VCVCC	<i>arend</i>	/ˈæɹənd/	/əˈɹɛnd/	<i>irend</i>	/ˈɪɹənd/	/ɪˈɹɛnd/
			<i>anunt</i>	/ˈænənt/	/əˈnʌnt/	<i>inunt</i>	/ˈɪnənt/	/ɪˈnʌnt/
			<i>elont</i>	/ˈɛlənt/	/ɪˈlɑnt/	<i>ilont</i>	/ˈɪlənt/	/ɪˈlɑnt/
			<i>ebint</i>	/ˈɛbənt/	/ɪˈbɪnt/	<i>ibint</i>	/ˈɪbənt/	/ɪˈbɪnt/
C	VC	CVCVC	<i>detep</i>	/ˈdɛtɛp/	/dɪˈtɛp/	<i>dotep</i>	/ˈdɔtɛp/	/dɔˈtɛp/
			<i>denen</i>	/ˈdɛnən/	/dɪˈnɛn/	<i>donen</i>	/ˈdɔnən/	/dɔˈnɛn/
			<i>relom</i>	/ˈɹɛləm/	/ɹɪˈlɑm/	<i>rolom</i>	/ˈrɔləm/	/ɹɔˈlɑm/
			<i>retud</i>	/ˈɹɛtəd/	/ɹɪˈtʌd/	<i>rotud</i>	/ˈrɔtəd/	/ɹɔˈtʌd/
	VCC	CVCVCC	<i>denont</i>	/ˈdɛnənt/	/dɪˈnɑnt/	<i>donont</i>	/ˈdɔnənt/	/dɔˈnɑnt/
			<i>denund</i>	/ˈdɛnənd/	/dɪˈnʌnd/	<i>donund</i>	/ˈdɔnənd/	/dɔˈnʌnd/
			<i>rebint</i>	/ˈɹɛbənt/	/ɹɪˈbɪnt/	<i>robint</i>	/ˈrɔbənt/	/ɹɔˈbɪnt/
			<i>redond</i>	/ˈɹɛdənd/	/ɹɪˈdɑnd/	<i>rodond</i>	/ˈrɔdənd/	/ɹɔˈdɑnd/
CC	VC	CCVCVC	<i>probal</i>	/ˈpɹɪəbəl/	/pɹɪəˈbæl/	<i>prabal</i>	/ˈpɹɪəbəl/	/pɹɪəˈbəl/
			<i>pronip</i>	/ˈpɹɪənəp/	/pɹɪəˈnɪp/	<i>pranip</i>	/ˈpɹɪənəp/	/pɹɪəˈnɪp/
			<i>prelel</i>	/ˈpɹɪɛləl/	/pɹɪəˈlɛl/	<i>pralel</i>	/ˈpɹɪɛləl/	/pɹɪəˈlɛl/
			<i>prepem</i>	/ˈpɹɪɛpəm/	/pɹɪəˈpɛm/	<i>prapem</i>	/ˈpɹɪɛpəm/	/pɹɪəˈpɛm/
	VCC	CCVCVCC	<i>pronast</i>	/ˈpɹɪənəst/	/pɹɪəˈnæst/	<i>pranast</i>	/ˈpɹɪənəst/	/pɹɪəˈnæst/
			<i>provand</i>	/ˈpɹɪəvənd/	/pɹɪəˈvænd/	<i>pravand</i>	/ˈpɹɪəvənd/	/pɹɪəˈvænd/
			<i>prebund</i>	/ˈpɹɪɛbənd/	/pɹɪəˈbʌnd/	<i>prabund</i>	/ˈpɹɪɛbənd/	/pɹɪəˈbʌnd/
			<i>predont</i>	/ˈpɹɪɛdənt/	/pɹɪəˈdɑnt/	<i>pradont</i>	/ˈpɹɪɛdənt/	/pɹɪəˈdɑnt/

Table 17. Test words and their different pronunciations used in the study initially conducted for Youssef Fadel's master's thesis

The target words were embedded in a carrier sentence and were underlined. Those sentences were taken from the *Corpus of Contemporary American English*, so that the target words were placed in sentence-final position and were always preceded by an unstressed function word (e.g. *to*, *the*) so as to keep the rhythmic context constant as this was found to be relevant in previous studies (see e.g. Kelly & Bock 1988). The sentences were presented visually in random order. The participants were presented with two possible pronunciations of the target word and were asked to choose the one that they prefer, as shown in Figure 19.

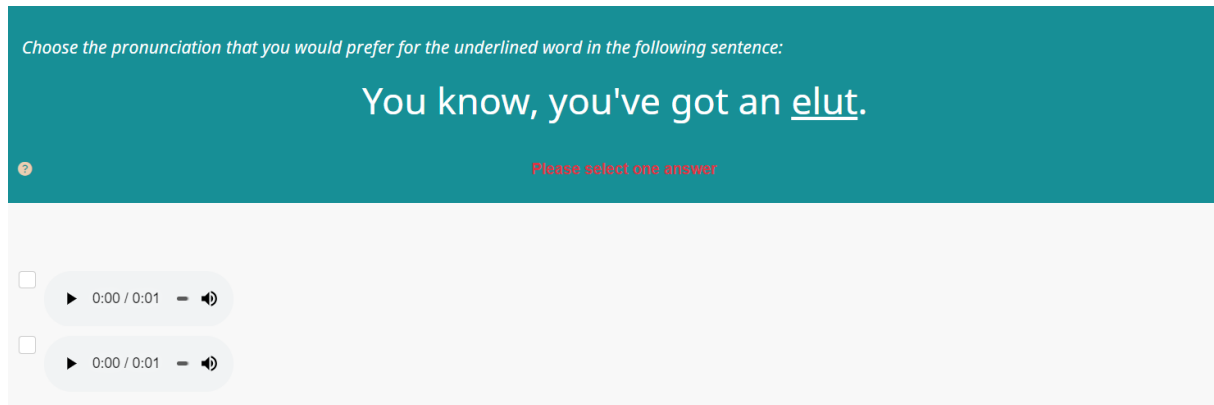


Figure 19. Illustration of the questionnaire used in the study

The participants were recruited by word of mouth, and the questionnaire also included sociodemographic information (age, gender, education, place of growing up). We were able to recruit 22 participants in that way, which were all native speakers of English. The results of that study were overall consistent with the literature, except that effects of prefixation only appeared when including participant as a variable in statistical models (Fadel 2023). One surprising result was the high rate of final stress that were preferred by the participants, as they chose the items with that stress pattern in 62% of cases. I will not detail those results but will focus on the replication that was conducted for the ERSaF project.

As the protocol was ready and the results were promising, we decided to replicate that study for the ERSaF project with more participants with a more homogenous language profile. This was done mainly along with Marie Gabillet, with whom I am currently writing a joint paper on that study. Those results were also presented in several conferences or workshops, including the 2024 OCP conference. We recruited 50 native speakers of British English using the platform Prolific (<https://www.prolific.com/>) who were paid for their participation to the study. This yielded 2,400 observations for the test words. The results were very similar to those found in the original study, still with a strong bias towards final stress: 64% in nouns and 70% in verbs. This is clearly higher than figures that can be found in the literature. For example, Hammond (1999) reports 20% of final stress in disyllabic nouns and 50% in disyllabic verbs. We ran a number of statistical models, and I will here focus on the results of a mixed regression model in which all the predictors were included, and participant was included as a random factor. The best model included an interaction between prefixation and syntactic category, and the effects of the model are shown in the plots in Figure 20. This plot shows the effects of each predictor while taking the other predictors in the model into consideration. For example, here Figure 20 shows that pseudowords which have a single coda consonant have final stress in about 63% of cases, while this figure increases to over 70% for those that have two coda consonants. In the raw data, those figures are actually 61% and 69%, respectively. The plot shows an adjusted version of the data which controls for the other variables that are present in the model. The 95% confidence interval is represented by the error bars.

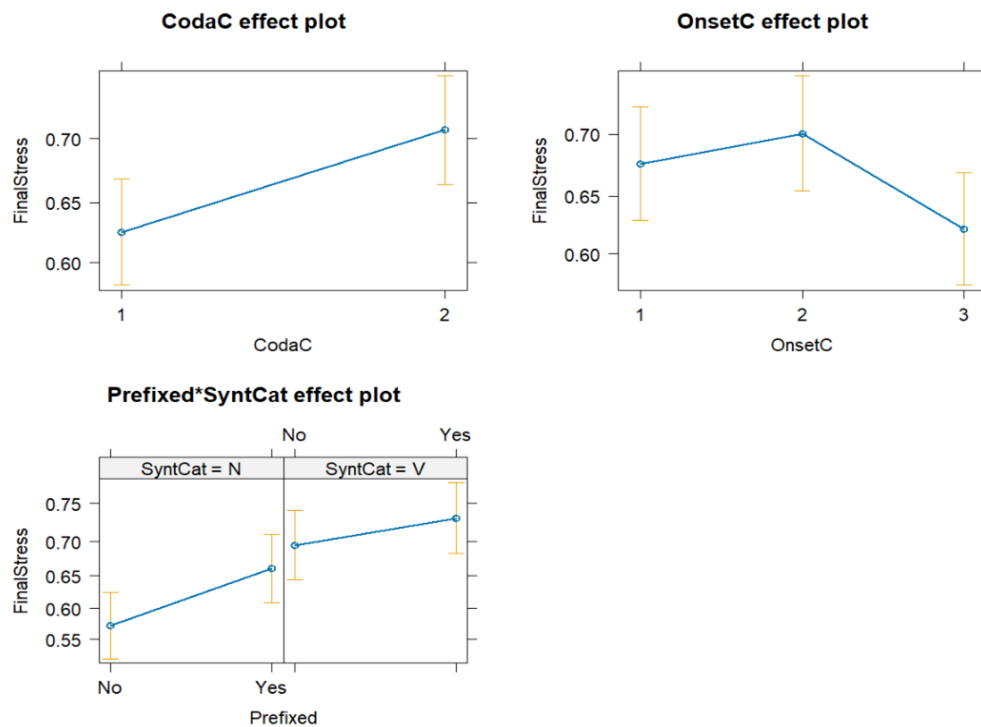


Figure 20. Effects plot for the best regression model for the preferred position of stress in disyllabic pseudowords

The first observation that can be made on those results is that they are overall consistent with the literature, but all the observed effects are very small, possibly because the strong predominance of final stress in all items weakens the effects of the factors that we tested. If we examine the effects of the different predictors, we first see that we find an effect of the number of coda consonants in the expected direction. What is interesting here is that this is also true of nouns, for which the literature usually assumes that the final syllable is extrametrical and thus should play no role in stress assignment. This result is consistent with those of previous pseudoword production studies (e.g. Domahs *et al.* 2014; Treiman *et al.* 2020). We also find an effect of the number of initial onset consonants, so that pseudowords with more onset consonants tend to have more initial stress, but the effect is not linear. As was mentioned above, we see here again that verbs expectedly have more final stress than nouns. Finally, the effect of prefixation goes in the expected direction, as pseudowords that have an initial string that can be taken to be a prefix have more final stress than the ones that do not. However, it was unexpected to find that the difference would be greater for nouns than for verbs (see §3.3.1.1). Once again, a possibility is that the rate of final stress for verbs is so high that it ‘hides’ the effects of prefixation.

Thus, the effects of the different predictors are overall consistent with the literature and are still found when all of them are taken into account at the same time. The most surprising aspect of the results is certainly the strong preference for pseudowords with final stress, and this calls for possible explanations. A first explanation could be that this is a task effect: somehow participants in a judgement study are more likely to choose final stressed pseudowords than if they were asked to produce the same pseudowords. One element supporting this are the results reported by Guion *et al.* (2003), who conducted both a production and a judgement study. Their results indeed show a greater proportion of final stress in the judgement study than in the production study for nouns. As for verbs, they have slightly lower rates of final stress in the

judgement study than in the production study, so that the difference between verbs and nouns is much smaller in the judgement study than in the production study. Moreover, Marie Gabillet got the data from Treiman *et al.* (2020) and extracted the items that had the same syllabic structures as our own stimuli (which, in part, were taken from their study, so some pseudowords were common to both studies). Note that those do not have potential prefixes as their initial strings except four pseudowords in <e->. If we leave those out, their participants produced final stress in 14% of nouns and 20% of verbs, which is far below the 57% and 68%, respectively, observed in our judgement data for non-prefixed pseudowords. A second option, which was suggested to us by Péter Szigetvári, is that the pseudowords were perceived as foreign, thus making them more likely to have final stress. While this may be part of the explanation, it cannot be the whole story. Indeed, in a recent study of stress in loanwords in English, P. Fournier (2024), finds only 12% of final stress in disyllabic loanwords taken from languages other than French, using dictionary data. As for French loanwords, the rate is much higher (around 50%, as we found in Dabouis & P. Fournier (2024), but not as high as that found in our study. Finally, a third explanation might have to do with the binary choice that was offered to participants. In this configuration, it has been shown that participants tend to equalize the number of responses for each option that is offered to them, and so it might explain why we find such a preference for final stress, even though the lexicon is more biased towards initial stress, especially for nouns. To conclude, let me just mention that close investigation of the data shows that not all prefixes behave in the same way, just as was found in the pseudo-verb study reported in §3.4.2.2. Therefore, we plan to investigate different characteristics of individual prefixes and how they might relate to the stress choices made by participants. Here too, those could include the proportion of words sharing the prefix and for which the base is free, the type and token frequency of the prefix, or the prefix likelihood measure developed by Wurm (1997, 2000). Preliminary tests with prefix likelihood suggest that this could indeed be a relevant factor.

3.4.5. *Disyllabic verb-noun pairs*

Let us conclude this section on stress and syntactic categories with an issue which is one of the first ones that I worked on: stress in prefixed disyllabic verb-noun pairs. This was the topic that the morphophonology team of the LLL was working on when I joined them in 2011. We made a series of presentations between 2012 and 2014, but never got around to publishing our study, possibly for lack of convincing results. As those studies are quite old, and because I would certainly not do them again in the same way, I will only briefly summarize what those consisted in. Then, I will turn to more recent explorations on the issue.

English notoriously has words for which primary stress seems to be contrastive as it may distinguish nouns and verbs such as *incrèase_N* and *incréase_V*. Within the Guierrian tradition, it has been found that, among disyllabic verb-noun pairs, this behaviour is almost exclusively restricted to opaque prefixed words (exceptions include pairs in *-ment* such as *tòrmènt_N* and *tòrmént_V* or isolated pairs such as *fìnànce_N* and *fínànce_V*), and that it is far from being the majority behaviour. For example, J.-M. Fournier (2010b) finds that there are about 200 such pairs and identifies the three types of stress behaviour in (55).

- (55) a. **Verbal** (oxytonic; $\approx 60\%$): both noun and verb are stressed on their second syllable (e.g. *concern, eclipse, mistake, supply*);
 b. **Alternating** (diatonic; $\approx 30\%$): the verb has final stress, and the noun has initial stress (e.g. *addict, import, permit, record*);
 c. **Nominal** (paroxytonic; $\approx 10\%$): both noun and verb are stressed on their first syllable (e.g. *contact, distance, profile, surface*).

A type that does not exist is a configuration in which the verb has initial stress and the noun final stress. This configuration has been found to be the only “unstable state” diachronically while the other three usually show stability through time (Sonderegger & Niyogi 2013). However, there had not been, and there still has not been, a full account of why certain pairs belong to a certain type and not another. This is the question that we set out to answer in 2011.

We started out from an extraction of all disyllabic verbs in the LLL’s Dictionary Database, with pronunciations from Jones (2006) and Wells (2008) for British and American English, and from the online Macquarie Dictionary for Australian English. We then applied a frequency filter on the data: we excluded a pair if one of its members had a token frequency of under 0.5 per million in the *Corpus of Contemporary American English (COCAE)*. This was an arbitrary decision, but we wanted to make sure that the items that we were studying were actually being used in both categories. The final dataset contains 186 pairs, and their distribution is nearly identical to that reported by J.-M. Fournier (2010) above. One striking observation that we made was the important intravarietal stress variation that we found in the data, with about a fifth of the pairs exhibiting stress variation in either the verb, the noun or both. This variation was highest in alternating pairs, which we interpreted as a sign that this is an unstable configuration. However, stress variation across different varieties of English is very limited, as had been previously found by M. Martin (2011). In our first presentation (Abasq *et al.* 2012), we gave a detailed account of that variation, and we tested one hypothesis on the possible determinants of stress profile of prefixed disyllabic verb-noun pairs: relative frequency. We tested whether a higher frequency of one member of the pair would favour the typical stress pattern of its category (e.g. if the noun is more frequent than the verb, we expect it to be more likely for the pair to be of the nominal type) and whether equivalent frequencies meant a greater likelihood for the pair to belong to the alternating type. This was done by calculating the ratio of frequencies based on the *COCAE*. Our results show limited effects of frequency, as can be seen from Figure 21, where we do find more nominal types if nouns are more frequent than verbs, but otherwise the distribution of the stress profiles does not seem to be affected by relative frequency.

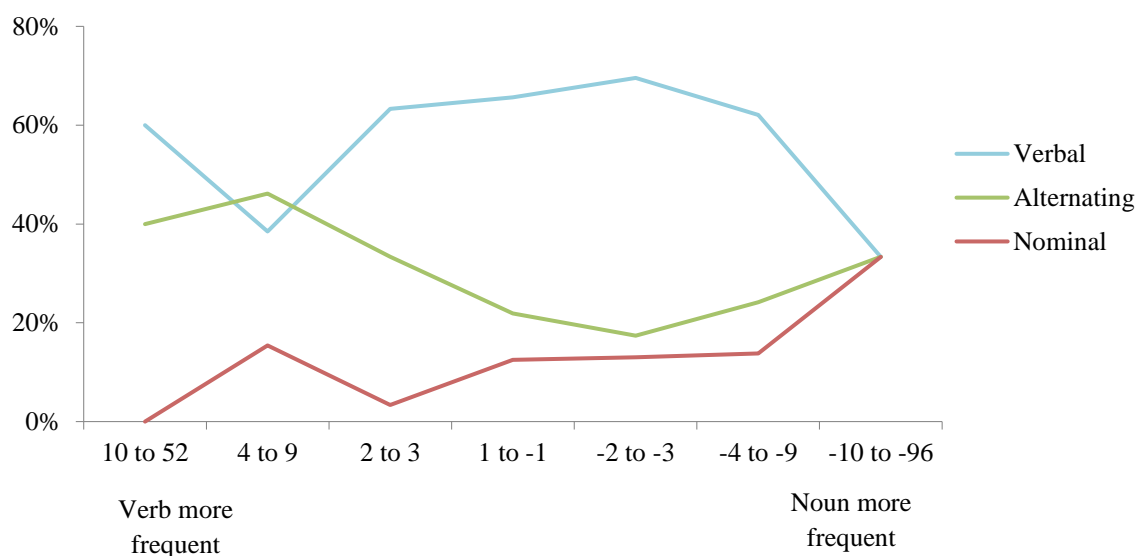


Figure 21. Stress profile of prefixed verb-noun pairs depending on their relative frequency in British English, based on Wells (2008)⁸²

In our second presentation (Abasq *et al.* 2012), we added an analysis of the vowels found in the unstressed syllable of words to see if it could predict the stress profile of the pair. Looking back at that study, we did not do it in a way that provides an answer to our question, as what we did was to see what kinds of vowels we could find in unstressed syllables in pairs of different stress profiles. We should probably have tried to neutralize vowel length and test if syllable structure can predict different stress profiles. Nonetheless, we made two interesting observations. First, we found a strong difference between first syllables and second syllables, which we interpreted as reflecting the different morphological natures of those syllables (prefix vs. root, respectively). Second, we found that there were fewer reduced vowels in those syllables in alternating pairs and in words that have a stress variant than in nominal or verbal pairs. As was seen in §3.3.1.1, this is used to argue that nouns that have the profile $\acute{\sigma}\grave{\sigma}$ inherit their secondary stress from the final-stressed verb that they are (supposedly) derived from through conversion.

In the follow up version of that study (Dabouis *et al.* 2014), we did do a more direct test of syllable weight, so that we were able to measure to what extent the weight of the final syllable correlates with the stress profile of the pair. Our results are shown in Figure 22, in which we can clearly see that lighter final syllables seem to favour the nominal type.

⁸² In order to represent ratios, we chose to turn ratios that would be inferior to 1 into negative ratios superior to 1 for better visualization.

THE INTERACTION BETWEEN MORPHOLOGY AND PHONOLOGY

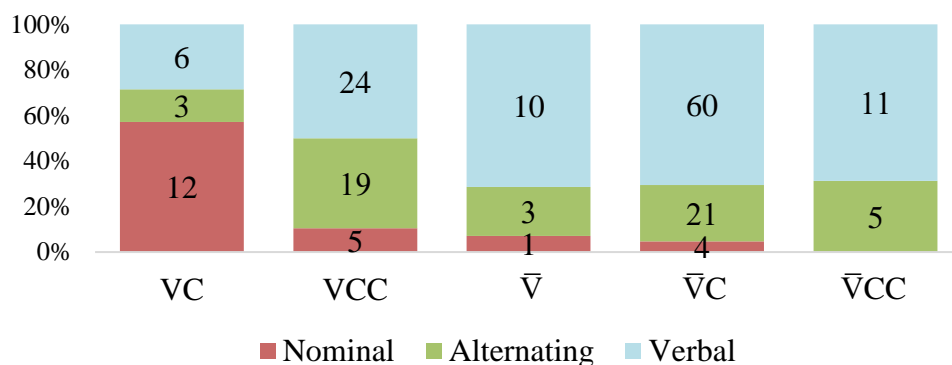


Figure 22. Stress profile of prefixed verb-noun pairs depending on the weight of their final syllable in British English

Another aspect of that study was a test of the hypothesis that diachronic conversion would still be reflected in synchronic stress patterns. Concretely, the prediction is that if the noun appeared earlier than the verb, we would expect the pair to be more likely to belong to the nominal type, if the verb appeared earlier than the verb, we would expect the pair to be more likely to belong to the verbal type, and if both members of the pair appeared at roughly the same time, we would expect the pair to be more likely to belong to the alternating type. Thus, we collected the dates of first appearance from the *OED* for both members of our pairs, and treated a member as having appeared earlier than the other if its first date of attestation is at least fifty years earlier than the first date of attestation of the other member. This yielded the results shown in Figure 23, which again show tendencies: it is more common to get the nominal type if the noun appeared before the verb.

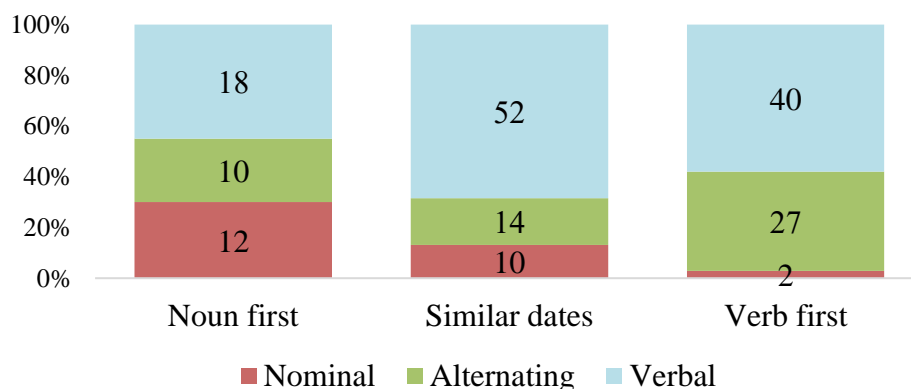


Figure 23. Stress profile of prefixed verb-noun pairs depending on their date of first attestation in the *OED*

Throughout these studies, we did not find any clear-cut generalizations that would neatly explain any of the patterns. In our 2014 poster, we merely hypothesized that the dominance of the verbal type can be explained by the fact that most disyllabic prefixed words are verbs, and that they are very regularly stressed on their second syllable, setting up a sort of ‘model’.⁸³ Our

⁸³ I had also run a number of tests on other potential variables, such as the prefix involved or the number of medial consonants, all of them showing the same kind of tendencies as those discussed in this section so far.

analyses had two crucial flaws. First, we did not conduct any multifactorial analyses, so that our results do not allow us to isolate the effects of each factor. For example, it is possible that we get more nominal type pairs if the nouns appeared earlier than the verbs because those nouns – for some reason – tend to have light final syllables (or the reverse). Second, we were looking for a categorical generalization that would explain the distribution of stress profiles. However, the results that we found were not showing any effects of that sort, and so this is probably what led us to abandon this project altogether and never attempt to publish our results. Now, I am much more comfortable with the idea that some processes are probabilistic so that such results, should the methodology be consolidated, would not be an issue for me today.

In 2020, I went back to those data with a different approach. The idea was to study the position of primary stress in disyllabic verbs which have an identified (near-)homographic noun and how the semantic relationship with the noun interacts with other factors in determining stress placement in verbs. Our dataset contains 1,746 such verbs, among which 1,255 were identified as denominal conversions in Dabouis & J.-M. Fournier (2023). In that chapter, we had not actually checked if the items identified as denominal conversions were stressed differently from the non-converted items. In a presentation made in the Laboratoire de Recherche sur le Langage’s (LRL) seminar (Dabouis 2020a), I explored the effects of the weight of the final syllable (as defined in Table 15), of the type of morphological structure, of the dates of first attestation, of the relative frequency of the noun and the verb and of the semantic dependency between the noun and the verb. The distribution of stress patterns found per morphological category is shown in Figure 24.

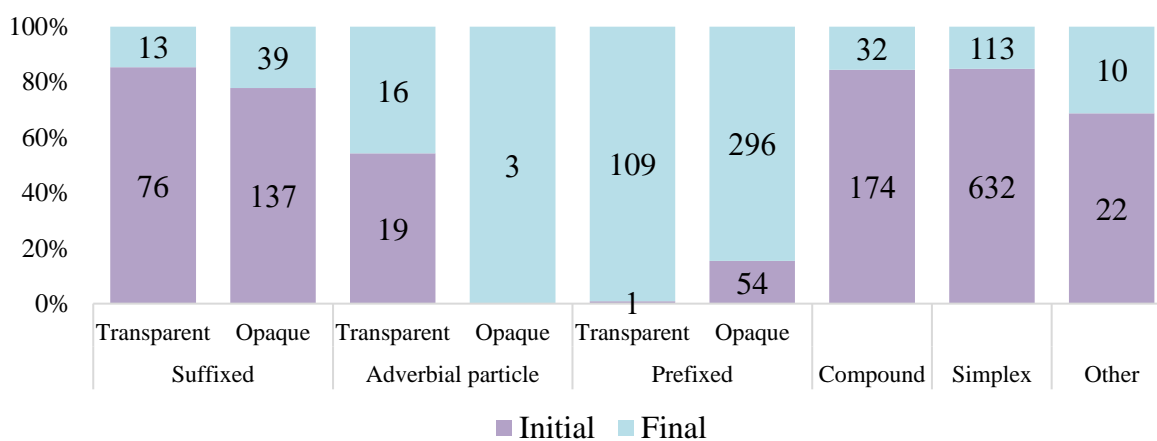


Figure 24. Position of primary stress in disyllabic verbs which have a homographic noun depending on different types of morphological structures

As can be seen in Figure 24, different morphological categories differ greatly in where primary stress falls, with final stress being the dominant pattern only in prefixed words and opaque verbs with an initial adverbial particle. An initial exploration of the data showed that only opaque prefixed words and words with an initial adverbial particle were significantly affected by predictors which relate to the relationship with the homographic noun (frequency, dates of attestation and semantic dependency). Therefore, in the rest of the study, I focused on those categories along with simplex items, which served as a baseline. Opaque verbs with an initial

adverbial particle were left out because there are only three such items. Now revisiting that subset of the data, I have left out dates of first attestation as this is not a predictor that is relevant for a synchronic analysis, and I have analysed the data using conditional inference trees. The formula that was used is shown in (56) and the plot generated is shown in Figure 25.

(56) `plot(ctree(Stress ~ Fq + Sem + Morph + Weight, data = data))`

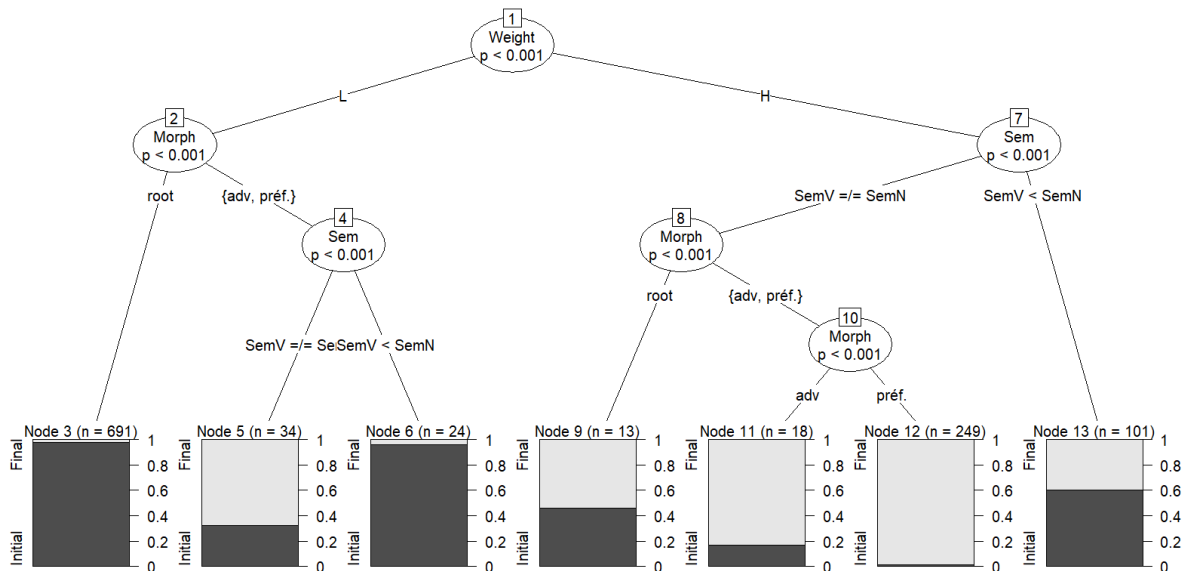


Figure 25. Conditional inference tree for position of primary stress in disyllabic verbs with a homographic noun. The predictors are morphological structure, weight of the final syllable and semantic dependency.

As can be seen in Figure 25, relative frequency does not turn out to be significant for the classification of the data.⁸⁴ The first significant split in the data is the weight of the final syllable, with overall more initial stress (in darker grey) if the final syllable is heavy than if it is light. Among words with a light final syllable, simplex words (coded as ‘root’) show near-categorical initial stress (e.g. *ámbush*, *chícken*, *gállop*, *máster*, *shéltér*). Prefixed words and words with an initial adverbial particle have the same behaviour if the meaning of the verb is dependent on that of the noun (e.g. *dístance*, *énvý*, *prófit*, *súrface*) but have final stress in 68% of cases if it does not (e.g. *attáck*, *expréss*, *incénse*, *outspréad*). Among items of that latter category which have initial stress, we find almost exclusively historically prefixed words for which the detection of that structure is synchronically probably very difficult as they have little to no paradigmatic support (e.g. *cóver*, *édít*, *óffer*, *rálly*). As for items with a heavy final syllable, simplex items whose meaning is not dependent on that of the noun differ from those with a heavy final syllable as we find an even distribution between initial stress (e.g. *chállenge*, *fóray*, *íssue*) and final stress (e.g. *caréen*, *patról*, *shampóo*). If there is no semantic dependency between the verb and noun and the final syllable is heavy, then stress is on the final syllable in prefixed verbs and verbs with an initial adverbial particle (e.g. *abúse*, *decáý*, *explóit*, *repróach*, *uplóad*). If the semantics of the verb are dependent upon those of the noun and the final syllable

⁸⁴ Note however that it was here measured as a binary variable with either the noun being more frequent than the verb using the *OED*’s frequency bands, or the noun being as frequent or less frequent than the verb.

is heavy, then there is a relatively even distribution between initial (e.g. *báckground, déluge, éxile, hárvest, témpest*) and final stress (e.g. *accént, cartóon, eclípse, machíne*), with a small advantage for the former, regardless of morphology. So, we can conclude from those observations that initial stress is indeed favoured in cases in which the verb is semantically dependent on its homographic noun, independently of the other factors.

This small exploration suggests that we look at semantic dependency more closely in future investigations of disyllabic verb-noun pairs. Indeed, there is diachronic evidence from Castanier (2023) that alternating pairs arise when new nominal meanings emerge that are not directly related to the meaning of the verb (e.g. *the act of V-ing, the fact of being V-en*).⁸⁵ It also shows the necessity of multifactorial analyses as the interaction between different factors may be quite complex. Finally, future studies of stress in disyllabic verb-noun pairs should also include factors that were not tested in our different studies, or not in a fine-grained enough fashion. This might include what prefix the word contains, more detailed semantic relationships, finer frequency measures and, for the dependent variable, the inclusion of stress variation. This is something that one of my master's students, Julien Penard, has been working for his master's thesis.

3.5. Suffix classes

The issue of suffix classes is an important topic in English morphophonology which has been the topic of numerous publications and debates between different theories. Focusing on stress for now, we find suffixes that are associated to a fixed stress pattern in the derivatives that contain them, which means that a derivative may have primary stress on a different syllable than its base (this is often called a “**stress shift**” as many theories have a derivational approach; see the examples in (57)).

(57)	Final	tráin ♦ tráinée	kítchen ♦ kitchenette
	Penult	héro ♦ heróic	áutumn ♦ àtúmnnal
	Antepenult	módern ♦ modérnity	húmid ♦ humidify

We also find classes of suffixed words that virtually always have stress on the same syllable as their base, sometimes leading to stress patterns that diverge from those of simplex words. Indeed, nouns with a closed penult tend to have penultimate stress more often than words with an open penult in simplex words (although this is partly conditioned by the type of vocabulary concerned; see Ch. 5) but this type of suffixed word systematically deviates from this if faithfulness to the base requires it (58a). Likewise, simplex words usually have their primary stress within a final three-syllable window, but this type of suffixed word may have pre-antepenultimate primary stress if faithfulness to the base requires it (58b).

⁸⁵ Castanier's study also provides elements that alternating pairs are not an end point in the stress evolution of disyllabic verb-noun pairs as proposed by Sherman (1975), but in line with previous studies by C.-Y. Chen (2014, 2017). C.-Y. Chen's analysis support analysis of verb-noun pairs as following a “migration path” in which such pairs evolve from the nominal type to the alternating type and finally to the verbal type. This is consistent with the hypothesis that the alternating type is actually an unstable state, which we proposed in Abasq *et al.* (2012).

(58)	a.	cáptaincy	♦	cáptain	b.	advénturousness	♦	advénture
		devélopment	♦	develop		àspirationally	♦	àspiration
		nóvelty	♦	nóvel		cháacterless	♦	cháacter
		párenthood	♦	párent		símilarly	♦	símilar
		wízardry	♦	wízard		váriableness	♦	váriable

Finally, there are suffixes which oscillate between the two behaviours, sometimes ‘shifting’ stress (59a), sometimes not (59b).⁸⁶

(59)	a.	cómparable	♦	compáre	b.	compárable	♦	compáre
		vólcanist	♦	volcáno		deféatist	♦	deféat
		ímmunize	♦	immúne		américanize	♦	América
		nècessárilý	♦	nécessary		háazardously	♦	háazardous

This is actually the first topic I have worked on for my master’s thesis (Dabouis 2012) and the paper that came out of it (Dabouis 2016a), which both focused on the stress and vocalic behaviour of words containing the adjectival suffix *-al*. I was very inspired by P. Fournier's (2011) work on suffixal derivatives in *-ous* in the way I approached the literature review, and I reviewed the generative literature regarding this suffix, and suffix classes in general. Thus, this is how I first became acquainted with *SPE*’s analyses, Metrical Phonology and Prosodic Phonology. Then, as one of the aims of my dissertation was to clarify the theoretical foundations of the Guierrian School and how it relates to other existing theories, I dedicated a section to the issue of suffix classes. I started working on it again in 2020 with Marjolaine Martin, with the goal of doing a large overview of the literature on the issue and to test different claims that have been made on six suffixes whose behaviour is unstable (*-able, -atory, arily, -ent, -ive, -ize*) but, because of the pandemic, we did not make as much progress as we would have liked. However, we did give a presentation on the theoretical overview on the occasion of a workshop for the forty years of the publication of Lionel Guierre’s dissertation (Dabouis & M. Martin 2022).⁸⁷ We intend to publish this overview and to continue the study of the six suffixes. In the rest of this section, I draft this overview and briefly present how different approaches deal with what I see as being two of the main controversies, shown in (60).

- (60) a. Is the difference between stress-shifting affixes and stress-preserving affixes **procedural** or **representational**?
- b. Is the difference between stress-shifting affixes and stress-preserving affixes **arbitrary** or **motivated**?

There is a wide array of answers to those questions, sometimes offering a mix of both options for each of those two key choices.

⁸⁶ The different suffixes that appear in (59) do not constitute a uniform group, as will be seen in the sections that follow.

⁸⁷ For that occasion, we also worked together to edit a new version of Guierre’s dissertation that was made available to the members of the ALOES (Association des anglicistes pour les études de langue orale dans l’enseignement supérieur, secondaire et primaire).

3.5.1. SPE's *boundaries*

In order to account for the different stress behaviours associated with different affixes, the authors of *SPE*'s propose a system of **boundaries** which are idiosyncratically associated with different affixes. Those boundaries are inserted into segmental strings but have the feature [-segment]. For suffixes, two boundaries are used:

- #: This is the word boundary which is inserted at the beginning and end of strings dominated by a major category. All suffixes are separated from their base by this boundary as a default. It is used to capture the phonological behaviour of neutral suffixes (e.g. #*less*, #*ness*, #*ly*), as it is visible to phonological rules and may thus block their application. Concerning stress, its effect is to block the application of the Main Stress Rule, so that this rule only applies to strings that do not contain such a boundary (*SPE*: 85). It is assumed to be “syntactic”, in that # is not stored in the lexicon and it represents boundaries that are computed on-line;
- +: This boundary is invisible to phonological rules, which means that complex words which contain this boundary undergo the same rules as simplex words. By language-specific convention, it is assumed that the # is changed to + before certain suffixes (e.g. +*ity*, +*ify*, +*ic*).⁸⁸ It is assumed to be “morphological”.

SPE's model is a **cyclic model**, and so even though complex words with suffixes introduced by + obey the same rules as simplex words, phonological differences between them may arise as a consequence of cyclicity. For example, the suffix *-(at)ion* comes with a + boundary and may trigger a reassignment of primary stress further to the right, but stresses assigned during previous cycles may be retained. If we represent the idea without using *SPE*'s formalism, the process can be illustrated as in (61).

(61)	condense + ation	compensate + ion
First cycle	cond/é/nse	cóm̩p/ə/nsàte
Second cycle	cònd/ε/nsátion	còm̩p/ə/nsátion

In both cases, the introduction of +(at)ion triggers a rightwards shift of primary stress, but the difference between the two words can be explained by the fact that the second syllable of *condensation* has received primary stress during the first cycle, but not that of *compensation*.⁸⁹

The authors note (*SPE*: 85-86) that the phonological manifestations of # go beyond stress, as there are two processes that occur word-finally which also occur before stress-preserving suffixes:

- /g/-deletion after a nasal (e.g. *sing#ing* /'sɪŋɪŋ/, *sing#er* /'sɪŋə/, *king#ly* /'kɪŋlɪj/, *king#let* /'kɪŋlət/) while this does not occur word-medially in the absence of such suffixes (e.g. *mingle* /'mɪŋgəl/, *finger* /'fɪŋgə/);

⁸⁸ As noted by Renate Raffelsiefen (p.c.), *SPE*'s analysis predicts that cases with bound roots and consonant-initial suffixes (e.g. *grateful*, *ointment*, *gormless*) are predicted to behave like words with no morphology. As will be seen in §3.5.5.1, she argues that this is not the case.

⁸⁹ In Wells (2008) this is a matter of variant orderings, as *condensation* is given as /ε ~ ə/ while *compensation* as /ə ~ ε/.

- syllabicity of sonorants, so that they can remain syllabic before a suffix introduced by #. They illustrate this with the pair *twinkling* ['twɪŋklɪŋ] ('instant', analysed as *twinkle+ling*) and *twinkling* ['twɪŋklɪŋ] (participle of *twinkle*, analysed as *twinkle#ing*).

However, apart from a note on *-ly* (*SPE*: 142), which is generally assumed to be preceded by # but may occasionally shift primary stress rightwards to avoid long strings of unstressed syllables while usually neutral (e.g. *ordinarily* ♦ *ordinary*), they do not mention the possibility that an affix may be associated with two different boundaries.⁹⁰

Therefore, *SPE* accounts for the difference between different suffix classes using only a **representational** device (boundaries), although the cycle is used to capture some of the phonological characteristics of suffixed words containing + boundaries procedurally. The choice of a boundary is assumed to be entirely idiosyncratic, and so **arbitrary**.

3.5.2. The Guierrian approach

Let us take a digression from the mainstream generative tradition into the Guierrian approach, which may be considered a post-generative model. In his dissertation, Guierre (1979) adopts the formalism of *SPE* and its boundaries, but not the cycle. He proposes a detailed quantitative analysis of the stress patterns found in suffixed words based on a large dataset built using data from Jones (1963). Guierre preserves the distinction between **neutral suffixes** and what he calls **strong “endings”**, which may or may not be suffixal. In doing so, he takes over Halle & Keyser’s (1971) observation that certain endings have a stress behaviour that is uniform, regardless of their morphemic status (e.g. final primary stress in words which end in two identical orthographic consonants and a final silent <e>, which includes the suffix *-ette*, e.g. *giraffe*, *kitchenette*). He also notes that ordering matters and that the rightmost strong ending is the one that prevails in imposing its preferential stress pattern.

He also adds an additional category, with an associated boundary †, which he calls “**non strictly neutral suffixes**” (Guierre 1979: 605) which have the properties in (62), which is my own translation and adaptation.

(62) **Non-strictly neutral suffixes**

- a. are not neutral when they are
 - i. juxtaposed to a base with final stress: *narrative* (♦ *narráte*)
 - ii. substituted to a base with penultimate stress: *volcanist* (♦ *vòlcáno*)
- b. are neutral in all other cases.

He notes that their orthographic forms are all either <-VCe>, <-VC₂(e)> or <-y> (*ibid.*: 639; e.g. *-ite*, *-ate_N*, *-ist*, *-age*, *-able*, *-ive*, *-ent*). However, this additional boundary is purely a notational device, and no formalization of what it does to stress computation is proposed.

⁹⁰ Within *SPE*’s framework, Aronoff (1976) does so for *-able*, as he assumes that there are two suffixes: *+able* and *#able*.

The use of boundaries for affixes has not been taken over by Guierre's successors, and the nature of what strong endings were doing was no longer attributed to those boundaries. Instead, J.-M. Fournier (1998) proposed that the distinction between the two classes of suffixes/endings is one that has to do with "modes of computation". He argues that neutral suffixes trigger "**computation by reference**", i.e. all they do is to refer to the pronunciation of the base, while strong endings block the reference to the base and trigger a "**direct computation**" of the pronunciation of the derivative. He notes that this direct computation may trigger stress shifts of the kind in (57), but may also trigger vowel shifts such as *div/áj/ne* ♦ *div/í/nity*. He calls this direct computation because the words that contain strong endings quite generally have stress patterns and vowels that conform to those found in simplex words. There are only a few endings for which one needs to assume that they do more than trigger direct computation (e.g. *-ic*, *-C₂ate*, all endings imposing final stress).⁹¹

Fournier also notes that, although non-suffixal endings may have the same overall behaviour as suffixes, suffixed words more strongly obey the generalizations associated to an ending than non-suffixed words. For example, words in <-ee>/-lj/ quite generally have final primary stress, but that is more often true of suffixed words (e.g. *adòptée*, *èndorsée*, *hònorée*) than of non-suffixed words (e.g. *cóffee*, *lichée* ~ *lichee*, *tóffee*).

As he notes in J.-M. Fournier (1994), direct computation actually concerns primary stress only, but not pretonic secondary stresses and the secondary stressed vowels. He shows that, even when the pattern of secondary stress is the same as that commonly found in simplex words, $\sigma\sigma(-)$, the secondary stressed vowel may diverge from what he calls "structural rules" (i.e. the generalizations that apply in simplex words) so as to be faithful to the vowel found in the base (e.g. *m/èj/jorétte* ♦ *m/éj/jor*). More generally, everything that is found left of primary stress tends to be faithful to the base, as long as being faithful does not generate stress clashes. This is something on which I have worked and will discuss in more detail in §4.5.1.

More recently, Castanier (2016: 693) suggests a new type of neutral suffix: **autostressed neutral suffixes**. The only case that is discussed is *-ee*, which takes primary stress and often does not trigger any modifications in the base except subsidiarization of primary stress to secondary stress (e.g. *parole* /pə'ɹəwɪ/ ♦ *parolee* /pə'ɹəwɪ'lj/). This kind of suffix would resemble semantically transparent prefixes, which bear secondary stress while their base bears primary stress, except that, in this type of suffixed words, the prominence between base and affix is reversed.

Finally, both Castanier and Trevian (2007, 2015) note that long sequences of post-tonic syllables seem to be increasingly dispreferred, so that stress shifts occur in a number of derivatives with what are usually assumed to be neutral suffixes: *-arily*, *-orily*, *-√Cable* and *-atory* all increasingly adopt antepenultimate primary stress, as shown in (63).⁹²

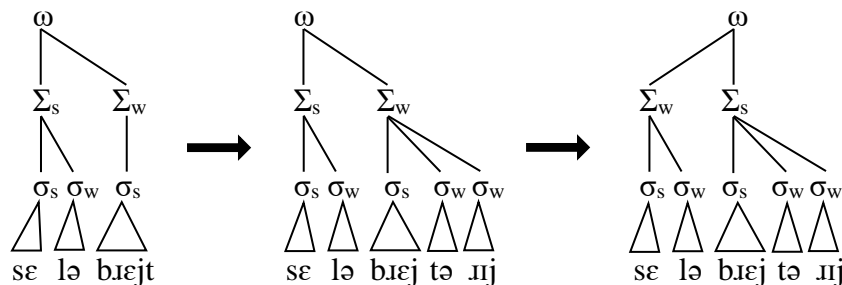
⁹¹ Observing the fact that opaque prefixes do not appear to have any effect in verbs in *-ate* (e.g. *démonstrâte*, *intonâte*, *renovâte*), J.-M. Fournier (1998) proposes that strong endings impose a form of morphological fusion, so that they block the interpretation of morphological structure beyond that of the suffix.

⁹² This is consistent with *SPE*'s observation on *-ly* that was reported in the previous section. In the case of *-orily*, the change mainly affects American English.

(63)	Base	Old pronunciation	New pronunciation
	<i>nécessary</i>	<i>nécessarily</i>	<i>nècessàrily</i>
	<i>obligatory</i>	<i>obligatorily</i>	<i>oblìgatorìly</i>
	<i>réconnize</i>	<i>réconnizable</i>	<i>rèconnìzable</i>
	<i>cèlebrate</i>	<i>cèlebratory</i>	<i>cèlebràtory</i>

On top of providing a detailed analysis of the diachronic changes that have taken place to lead to those new pronunciations (see §3.5.7 on *-atory*), Castanier notes that this can be easily accounted for in Metrical Phonology (Hayes 1981): there is often a post-tonic secondary stress, and so adding a suffix makes the final foot heavier, which eventually leads to a shift in prominence, as illustrated in (64), which is adapted from Castanier (2016: 593-594).

(64) **Evolution of stress in *celebratory***



Thus, although the Guierrian approach does not embrace a modular framework, where Guierre took over the **representational** device of, Fournier uses a **procedural** device akin to the cycle: each addition of a strong suffix triggers an application the (grapho)phonological grammar, but the addition of a neutral ending does not. However, the class to which each suffix belongs is assumed to be entirely **arbitrary**, although related to history: most neutral suffixes are Germanic, and most strong endings are Latinate (J.-M. Fournier 2007). More recently, J.-M. Fournier (2024) proposes that the difference between different types of suffixes mainly has to do with the fact that the words that contain different types of suffixes belong to different parts of the lexicon. This means that the difference is no longer viewed as a morphosyntax-phonology interaction but rather that it reflects different subgrammars corresponding to different sublexica. In the terms that will be seen in Ch. 5, the distinction concerns §CORE-NATIVE (neutral affixes) versus the other sublexica.

3.5.3. Stratal models

Shortly after the publication of *SPE*, Siegel (1974) proposes a more in-depth analyses of affix classes, in which affixes introduced by + are called “Class 1” affixes and those introduced by # “Class 2” affixes, a terminology that is still widely used. She also introduces the idea of a **level-ordered morphology**, so that Class 1 affixation precedes Class 2 affixation. Moreover, she assumes that Class 1 affixation precedes cyclic stress assignment rules while Class 2 affixation follow them. Those ideas were at the centre of Lexical Phonology, a theory which sought to formalize the interaction between morphosyntax and phonology using strata, the structure of which is shown in Figure 26.

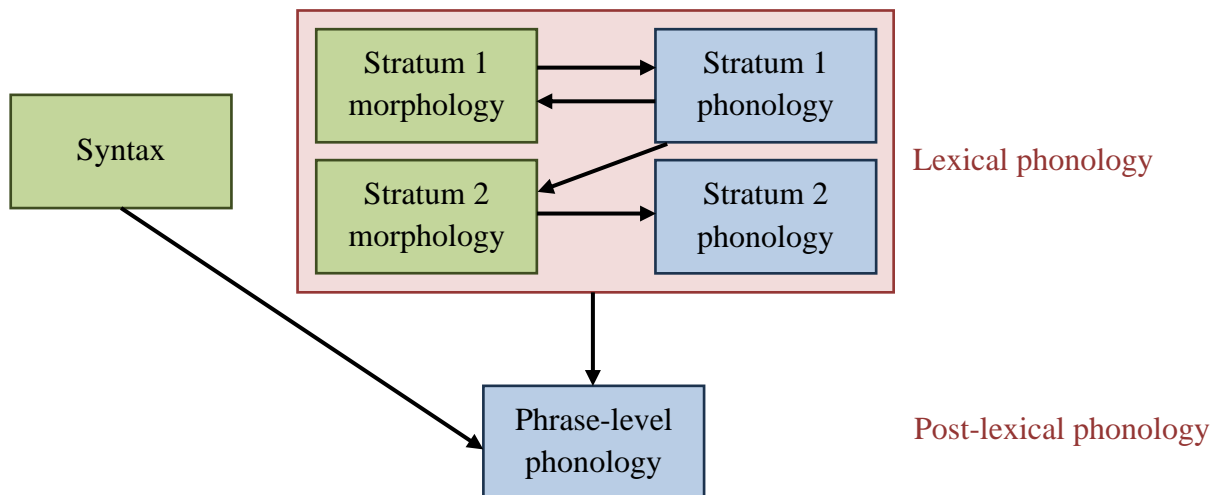


Figure 26. The organization of grammar in Lexical Phonology

One crucial aspect of the theory is that, like its more recent successor Stratal Phonology, each stratum has a different phonological grammar (Bermúdez-Otero 2018b calls them \mathcal{F} -functions). In Lexical Phonology, affixes are arbitrarily assigned to a stratum, and so they have sometimes been called “Level 1/2” affixes or “Stratum 1/2” affixes. The first stratum in those models is usually assumed to be cyclic, while this is not necessarily the case of later strata, so that stress computation occurs at each concatenation of a Class 1 affix, and this can be used to capture what *SPE* did with the cycle.

An important contribution of the Lexical Phonology literature is the extension of the properties of those affix classes beyond stress, /g/-deletion and syllabic consonants. For example, Kiparsky (1982) notes that the rule of Trisyllabic Shortening applies to Class 1 derivatives (65a) but not to Class 2 derivatives (65b).

- (65) a. *opacity* /əw'pʌsətɪj/ ♦ *opaque* /əw'pɛjk/ b. *mightily* /'maɪtɪli/ ♦ *mighty* /'maɪtɪ/
- declarative* /dɪ'klaɪətɪv/ ♦ *declare* /dɪ'kleɪ:/ *bravery* /'brɛɪvənɪj/ ♦ *brave* /'brɛɪv/

Borowsky (1993) and J. Harris (1990) also describe a number of phonological processes found in diverse varieties of English which misapply in cases of Class 2 affixation. Cases of overapplication of processes in Class 2 affixation are summarized in Table 18.

Regular application in base	Regular non-application in Class 1 affixation /simplexes	Overapplication in Class 2 affixation
i. æ-tensing in closed syllable (New York City & Philadelphia) – B93		
<i>class</i> [klɛs]	<i>classic</i> [klæ.sɪk]	<i>classy</i> [klɛ.sɪ]
<i>mass</i> [mɛs]	<i>massive</i> [mæ.sɪv]	<i>massable</i> [mɛ.səbl]
ii. Word-final THOUGHT/NORTH centralization (London Vernacular English) – H90 & B93		
<i>paw</i> [pɔə]	<i>pause</i> [pouz]	<i>paws</i> [pɔəz]
<i>boar</i> [bɔə]	<i>board</i> [boud]	<i>bored</i> [bɔəd]
iii. Word-final FACE monophthongization and lowering (Northern Irish English) – H90 & B93		
<i>play</i> [plɛ:]	<i>fate</i> [fɪət]	<i>playful</i> [plɛ:fəl]
<i>day</i> [dɛ:]	<i>vain</i> [vɪən]	<i>days</i> [dɛ:z]
iv. Aitken’s law: word-final lengthening (Scottish English) – B93		
<i>brew</i> [brɛ:]	<i>brood</i> [brʊd]	<i>brewed</i> [brɛ:d]
<i>knee</i> [ni:]	<i>need</i> [nid]	<i>kneed</i> [ni:d]
v. Rounding before coda [ɨ] (Adelaide Dialect) – B93		
<i>fool</i> [fu:ɨ]	<i>Julie</i> [jʊ.liy]	<i>fooling</i> [fu:.lɪŋ]
<i>boal</i> [bɔuɨ]	<i>polar</i> [pʌɯ.lɔɐ]	<i>bowler</i> [bɔu.lɔɐ]
vi. Cluster simplification (Standard English) – B93		
<i>damn</i> [dam]	<i>damnify</i> [damnəfɪj]	<i>damning</i> [damɪŋ]
<i>long</i> [lɒŋ]	<i>elongate</i> [ɪlɒŋgɛjt]	<i>longing</i> [lɒŋɪŋ]

Table 18. Cases of overapplication of phonological processes induced by Class 2 affixation (H90 indicates that the process is described in J. Harris (1990) and B93 that it is described in Borowsky (1993))⁹³

All the processes in Table 18 apply in certain syllabic conditions (e.g. closed syllables, word-final syllables), which may be modified by affixation. In all of these cases, the processes fail to apply in simplex words and in Class 1 derivatives when their syllabic conditions are not met, but they overapply in Class 2 derivatives. In Lexical Phonology, this is analysed by assuming that those processes apply before Class 2 affixes are added.⁹⁴ J. Harris and Borowsky also discuss one case of underapplication, shown in (66).

⁹³ The transcriptions are taken from the source articles, except in (vi) where no transcriptions were provided, and so those examples are transcribed using the standard used throughout this document.

⁹⁴ For example, Borowsky (1993) assumes a rule-based framework and posits that those rules apply at the word-level and that word-level (Class 2) affixes are added after. In Stratal Phonology, such processes are assumed to take place at the stem-level as word-level affixes are added before word-level phonology applies. Such differences in the stratal affiliation of the phonological processes whose domain excludes word-level affixes are discussed for example by Bermúdez-Otero & McMahon (2006) and I have discussed this issue in my review of Jensen (2022) (Dabouis 2022).

(66) **Northern Irish Dentalization** /t, d, n, l/ → [t̪, d̪, n̪, l̪] / ___ [θ, (ə)r]

- a. ma[t̪]er b. sani[t̪]ary c. fa[t̪]er
 a[n̪]them eleme[n̪]ary di[n̪]er

In that case, the process of dentalization requires the target consonants to be adjacent to the triggers of dentalization. The process correctly applies in simplex words (66a) and in Class 1 derivatives (66b), but fails to apply in Class 2 derivatives (66c). Here too, it can be assumed that the process takes place at Stratum 1 and thus cannot be affected by subsequent affixation.

In early generative phonology, those classes of affixes have been argued to have properties that go beyond phonology, as summarized by Raffelsiefen (1999) in Table 19.

properties:	Class I affixes:	Class II affixes:
morphological:	<ul style="list-style-type: none"> ◆ Can attach to non-word bases ◆ Unproductive 	<ul style="list-style-type: none"> ◆ Attach to words only ◆ Productive
Phonological:	<ul style="list-style-type: none"> ◆ Stress-determining ◆ Trigger Nasal Assimilation ◆ Trigger Trisyllabic Laxing 	<ul style="list-style-type: none"> ◆ Stress-neutral ◆ Block Nasal Assimilation ◆ Block Trisyllabic Laxing
semantic:	<ul style="list-style-type: none"> ◆ Yield idiosyncratic meaning 	<ul style="list-style-type: none"> ◆ Yield compositional meaning

Table 19. The properties of Class I and Class II affixes in early generative phonology according to Raffelsiefen (1999)

In that literature, we find observations that certain suffixes do not properly fit any of the two classes and sometimes show phonological behaviours of one class, and sometimes those of the other (e.g. *-able*, *-ize*, *-ment*; Selkirk 1982; Szpyra 1989). For example, *-ment* would be a Class 1 affix in constructions with bound roots such as *ornament* but a Class 2 affix when attached to a free base (e.g. *employment*).

As can be seen in Table 19, it has long been noticed that the bases to which the different types of affixes attach may be of different natures and that, more broadly, the different strata refer to different types of morphosyntactic units, hence the names now commonly used in Stratal Phonology: stem-level, word-level and phrase-level. As was seen in §3.2.2, Bermúdez-Otero (2018b) gives four key generalizations that establish which type of morphosyntactic unit define domains for which stratum, repeated in (67).

- (67) a. Roots do not define cyclic domains.⁹⁵
 b. Some stems define cyclic domains for the stem-level phonology (\mathcal{P}_{SL}).
 c. Words define cyclic domains for the word-level phonology (\mathcal{P}_{WL}).
 d. Utterances define cyclic domains for the phrase-level phonology (\mathcal{P}_{PL}).

⁹⁵ Some possible counterevidence to this claim will be presented in §3.5.5.1.

The central role of the base in stratification is exposed by Giegerich (1999), who argues that all affixes are “dual-level” and so that stratification cannot be driven by affix classes but by the **nature of the base**. The consequence is that being ‘Level 1’ or ‘Level 2’ is no longer a property of affixes but of words, which leads him to expose a series of criteria to identify the stratum to which a given construction belongs. Those are shown in (68).

- (68) a. **Nature of the base:** if an affix attaches to a bound root, then the construction is Stratum 1 (e.g. *final*, *social*, *lethal*). Otherwise, it is not necessarily Stratum 2 (e.g. *baptismal*, *hormonal*).
- b. **Productivity and semantic uniformity:** Stratum 1 processes are less productive than those of Stratum 2 and Stratum 1 constructions may be semantically non-compositional.
- c. **Stress behaviour:** a stress shift within the base indicates a Stratum 1 status. Giegerich notes that this is probably not true of autostressed suffixes (e.g. *-ee*, *-esque*).
- d. **Syllabification behaviour:** syllabic consonants cannot be the nucleus of a syllable in Stratum 1 constructions.
- e. **Phonotactic behaviour:** certain structures that are prohibited on Stratum 1 are licit on Stratum 2 (e.g. geminated consonants). This is the only criterion against Stratum 1 status.
- f. **Cyclic phonological rules and allomorphy:** Trisyllabic Shortening is inapplicable at Stratum 2, but its non-application does not necessarily indicate Stratum 2 status. Allomorphies in the base (e.g. *amplify* ♦ *amplification*, *permit* ♦ *permissible*) indicate Stratum 1 status.

More recent models such as Stratal Phonology are closer to the classical position of Lexical Phonology on the affiliation of affixes to strata: some are stem-level, others are word-level and yet others are dual-level. Occasionally, word-level affixes may occur in stem-level forms (e.g. see Bermúdez-Otero’s (2012) discussion of comparative *-er* or Bermúdez-Otero (2018b) on affix ordering).

Therefore, in stratal models, the behaviour of different affixes is captured **procedurally** through their different stratal affiliations, each stratum being associated with a different phonological grammar, and strata being assumed to be ordered. Moreover, some behaviours of derivatives containing stem-level affixes can be captured using stratum-internal cyclicity, as *SPE* did for the *condensation-compensation* cases (see also the discussion of Weak Preservation in §3.2.2). The stratal affiliation of affixes is assumed to be **mostly arbitrary** and, in some cases, it is assumed to be driven by the nature of bases.

3.5.4. OO-Correspondence

The approach to suffix classes in strongly parallel OT was initially proposed by Benua (1997), and her analyses are still being used today. As the model rejects the existence of strata, she proposes another device to capture differences between affixes: Class 1 and Class 2 affixes are indexed to different OO-Correspondence constraints, so that Class 2 affixes require a greater

faithfulness to the base than Class 1 affixes. The two basic constraints used in the analysis that she makes of primary stress placement and that are relevant here are shown in (69).

- (69) a. ALIGN((Hd)PrWd, R, PrWd, R)
 The head of the prosodic word is aligned at the right edge of the prosodic word (main stress is at the right).
- b. NONFINALITY
 Word-final syllables are not footed.

The constraints used for stress identity are called OO-ANCHOR, and they require correspondence of the edges and heads of feet. Benua's analysis is illustrated in (70) and (71), in which squared brackets represent the edges of feet.

(70) Recursion (A)

/obvious/	NONFINAL	OO ₂ -ANCHOR	ALIGN-R	>>
a. ob[ví.ous]	*!		*	
b. [ób]vi.ous			**	
☞ c. [ób]vi.ous			**	

Recursion (B)

>>	/obvious+ness/	NONFINAL	OO ₂ -ANCHOR	ALIGN-R
	a'. ob[ví.ous]ness			**
	b'. ob[ví.ous]ness		*!	**
☞	c'. [ób]vi.ous.ness			***

First, let us note that this is a more 'classical' OT model than the MaxEnt model that was seen in §3.4.2.1, in that it evaluates candidates on the basis of Strict Domination of constraints: higher-ranked constraints have absolute priority over lower-ranked constraints. In other words, violating one higher-ranked constraint is worse than violating any number of lower-ranked constraints. However, the model is special in that she uses assumes a "recursive" evaluation of the candidates, so that the constraint hierarchy is duplicated for each OO-Correspondence relationship between two strictly local words (i.e. separated by a single affix). Simpler words are given greater priority in recursion, so that a candidate that is eliminated in the recursion of a simpler word will be eliminated in further recursions. In (70), we can see that *obvious* gets assigned initial stress as candidate (70a) is eliminated as it violates NONFINALITY. Then, in the second recursion, the OO₂-ANCHOR constraint indexed to Class 2 affixes enforces stress identity with the candidate selected in the first recursion. Candidate (70a') is excluded altogether from Recursion (B) as its correspondent in Recursion (A) is eliminated. Then, the stress-shifted candidate (70b') is eliminated because it violates OO₂-ANCHOR, which leaves the stress-preserving candidate (70c').

The analysis for Class 1 affixes differs in that the OO₁-ANCHOR indexed to those affixes is ranked lower, so that markedness constraints may prevail and force stress further to the right in the derivative, as can be seen in (71).

(71) Recursion (A)

/origin/	NONFINAL	ALIGN-R	OO ₁ -ANCHOR	>>
a. o[rí.gin]	*!			
☞ b. [ó.ri]gin		*		
c. [ó.ri]gin		*		

Recursion (B)

>>	/origin+al/	NONFINAL	ALIGN-R	OO ₁ -ANCHOR	>>
	a'. o[rí.gi]nal		**		
☞	b'. o[rí.gi]nal		**	*	
	c'. [ó.ri]gi.nal		***!		

Here again, the first recursion assigns initial stress to the base (the only difference being that the foot is here disyllabic, but it is bimoraic as in (70)). However, in the second recursion, candidate (71c') violates ALIGN-R more times than candidate (71b'). Therefore, the latter is the optimal candidate, even if it violates OO₁-ANCHOR, which is ranked lower than ALIGN-R.

Benua uses the same theoretical machinery to account for the cases seen previously in Table 18. She argues against a purely representational solution which would consist in assuming that Class 2 affixes are not integrated in the phonological word of the base (e.g. Hammond 1999: 322-329; Aronoff & Sridhar 1983; Szpyra 1989: 178-200). She argues that those approaches cannot be right because there is evidence that (at least certain) Class 2 affixes syllabify with the base, as seen in case (v) of Table 18: in the Adelaide Dialect, Class 2 derivatives have a rounded vowel like their base (e.g. *fool* [fu:t] ♦ *fooling* [fu:.lɪŋ]) even though the triggering /l/ is no longer in the coda. This excludes an analysis such as *fooling* (fu:l)_ωɪŋ, because this would wrongly predict that the /l/ should be realized [ɫ], as this variety of English has [ɫ] in the coda and [l] in the onset.

The idea of having parallel cophonologies that would be specific to affix classes is still widespread today, some even pushing it further so that there are affix-specific cophonologies (see the next section for an example). Benua's approach is purely **procedural** although, as will be seen in the next section, this approach is sometimes mixed with representational solutions. As to arbitrariness, Benua (1997: 166-167) explicitly states that "Attempts to relate class membership to inherent properties of the affix, such as its etymology, morpho-syntactic features or prosodic shape, are unsuccessful" and that "Membership in a phonological class is **arbitrary**" (emphasis mine).

3.5.5. Two approaches to cohesion

3.5.5.1. Raffelsiefen

Taking over a proposal by Dixon (1977a, 1977b), Raffelsiefen (1999, 2004, 2005, 2015, 2023) proposes that the different behaviours found among English affixes are best analysed through a distinction between “**cohering**” and “**non-cohering**” affixes. Cohesion means that the affix fuses with the stress and syllabification domain of the stem, the phonological word. She argues that all vowel-initial suffixes are cohering while consonant-initial suffixes are all non-cohering, which means that they are prosodically separate from their stem.⁹⁶ Non-cohesion means that the affix is outside of the stress domain of the stem, which means that the phonological form of the affix may be one source of ‘stress-neutrality’. Other cases of ‘stress-neutrality’ found in cohering suffixes are attributed to OO-Correspondence constraints, that she calls “paradigm uniformity” constraints. In Raffelsiefen (2005), she illustrates this difference with examples that are morphologically similar but phonologically distinct, although all the suffixes involved are ‘stress-neutral’ (i.e. preserving): those in (72a) are ‘neutral’ because they are associated with high-ranking paradigm uniformity constraints requiring stress faithfulness to their base, while those in (72b) are ‘neutral’ because they are not integrated into the phonological word of the stem.

(72)	Morphological structure	Prosodic structure
a.	[[vínegar] _{STEM} [ish] _{HEAD SUFFIX}] _{WORD}	(vínegarish) _ω
	[[préside _{nt}] _{STEM} [y] _{HEAD SUFFIX}] _{WORD}	(préside _{nt} y) _ω
	[[cándidate] _{STEM} [ure] _{HEAD SUFFIX}] _{WORD}	(cándidature) _ω
	[[dífficult] _{STEM} [y] _{HEAD SUFFIX}] _{WORD}	(dífficult _y) _ω
b.	[[rádical] _{STEM} [ness] _{HEAD SUFFIX}] _{WORD}	((rádical) _ω ness) _ω
	[[cítizen] _{STEM} [ship] _{HEAD SUFFIX}] _{WORD}	((cítizen) _ω ship) _ω
	[[ínstant] _{STEM} [ly] _{HEAD SUFFIX}] _{WORD}	((ínstant) _ω ly) _ω
	[[néighbour] _{STEM} [hood] _{HEAD SUFFIX}] _{WORD}	((néighbour) _ω hood) _ω

As was seen in the previous section, there are arguments for rejecting that all stress-preserving suffixes prosodify separately from their stem, as the resulting prosodic structures make erroneous predictions (e.g. Aronoff & Sridhar (1983) propose to analyse *alphabetize* as (alphabet)_ωize, which wrongly predicts /t/-flapping in varieties that have it). The difference between cohering stress-shifting and stress-preserving suffixes is attributed to affix-specific cophonologies.

To support her claim, Raffelsiefen details a number of phenomena associated with cohering and non-cohering suffixation, some of which are briefly reviewed in (73).

⁹⁶ This is argued to be caused by some form of “deficiency” of the affix, such as the lack of nucleus or onset. Violating standard markedness constraints such as ONSET or NUCLEUS (which require a syllable to have an onset or a nucleus, respectively) is worse than violating the constraint requiring the right edge of the stem to be aligned with the right edge of a phonological word, and so such deficient affixes are argued to be integrated into the phonological word.

- (73) a. **Phonetic implementation as a window on prosodic structure:** allophonic realizations may be used as diagnostics on the prosodic integration of the suffix:
- i. aspiration and absence of flapping in the /t/ of *parrotize* suggests cohesion (cp. *parrot eyes* with no aspiration and possible flapping);⁹⁷
 - ii. further reduction of /ə/ before word-final sonorants: (*joy*)_ω*ful* ['dʒɔjʃəl]⁹⁸ vs. (*rifle*)_ω ['ɹɪʃlɪ] (this is based on transcriptions from Kenyon & Knott 1953);
 - iii. longer diphthong and more consonantal intervocalic consonant in non-cohering suffixation, e.g. (*shy*)_ω*ness* vs. (*minus*)_ω (Z. S. Harris 1951);
 - iv. misalignment of gestures between nasals and following voiceless fricatives, shown in pronunciation dictionaries as optional voiceless stops, e.g. *triumphal* /ˌtɹɪʃəm(p)ʃəl/. This is not found for words with non-cohering affixes, e.g. *shameful* /'ʃeɪmfəl/.⁹⁹
- b. **Perceived “perfect” rimes:** cohering affixations rime perfectly with other words that form a single phonological word (e.g. (*larval*)_ω /'lɑ:vəl/ = (*marvel*)_ω /'mɑ:vəl/), but this is not the case for non-cohering affixations (e.g. (*free*)_ω*ness* /'fri:nəs/ ≠ (*Venus*)_ω /'vi:nəs/);¹⁰⁰
- c. **Phonotactic restrictions on the right edge of the phonological word:** certain restrictions that hold for the right edge of simplex words also hold word-internally before non-cohering suffixes:
- i. ‘checked’ vowels are not allowed, and so stem-final ‘tense’ vowels may be preserved before non-cohering affixes (e.g. (*pity*)_ω*less* /'pɪtɪjləs/ vs. (*syphilis*)_ω /'sɪfɪləs/);¹⁰¹
 - ii. “Final-C effects”: certain consonantal structures are possible word-finally but not word internally: word-internal rimes followed by an onset (analysed as being followed by an empty nucleus), e.g. (*hemp*)_ω /'hɛmp/ cp. (*tempo*)_ω /'tɛmpəw/ may be found before non-cohering suffixes (e.g. (*large*)_ω*ness* /'lɑ:dʒnəs/). Likewise, it is possible to have an additional coronal obstruent word-finally (e.g. (*sphinx*)_ω /'sfɪŋks/) and before non-cohering suffixes (e.g. (*world*)_ω*ly* /'wɜ:ldli/);

⁹⁷ See also Bermúdez-Otero (2007b) on evidence from the allophonic behaviour of /l/ and /ɫ/ which also points to the integration of vowel-initial suffixes into the phonological word of their base.

⁹⁸ The transcriptions given in this section are my own as they do not affect the arguments put forward by Raffelsiefen, who uses transcriptions of American English.

⁹⁹ We can note that this is sometimes found in the kind of complex proper names that was seen in §3.3.2 (e.g. *Cromford* /'krɒmfəd/, *Livingston* /'lɪvɪŋstən/, *Stephenson* /'stɪjvənsən/) but not systematically. If this is confirmed by further investigation, it would support the analysis that such words are not simplex words.

¹⁰⁰ This is taken from Raffelsiefen (2024), in which no sources to support this claim are given.

¹⁰¹ As noted by Herment (2010), there is actually quite a lot of inconsistencies across different dictionary sources on the nature of the stem-final vowel before consonant-initial suffixes. She notes that Jones (2006) and Wells (2008) are consistent in showing /ɪ/ or /ə/ before *-ful* and *-ly*, but inconsistent for *-ment*, *-ness*, *-wise* and *-less*. This goes against Raffelsiefen’s (2024: fn. 20) suggestion that the presence of /ɪ/ “may indicate a sporadic fusion that affects individual common words, rather than properties of the suffixes *-ful* and *-ly*”. She justifies this by saying that there are words with those same suffixes that have /ɪj/ and cites *bountiful* and *sleepily*, but those two words have /ɪ/ or /ə/ in Jones (2006), Wells (2008) and the online *OED*, and I have not been able to find any cases with /ɪj/ before *-ful* or *-ly* in those sources. As pointed out by Raffelsiefen (p.c.), Upton & Kretzschmar (2017) do give /ɪj/ (transcribed /i/) before some words in *-iful*. It is unclear when they do so, as shown by the fact that *dutiful* is transcribed with /i/ while *undutiful* is transcribed with /ə/. Given the uncertainty arising from the dictionary data, it may be necessary to study this phenomenon using oral data.

- d. **Different properties in word formation:** in native word formation, only cohering affixation shows the following properties:
- i. stem allomorphy (e.g. stress shifts, *Vietnám+ése* → (*Vietnamése*)_ω vs. *ballóon+like* → (*ballóon*)_ω(*like*)_ω);
 - ii. stem selection (e.g. *emphas-is+ize* → *emphasize*, *amput-ate+ee* → *amputate* vs. *curious+ness*, *shovel+ful*)
- e. **Affix ordering:** non-cohering derivational suffixes cannot be followed by cohering ones (e.g. *((*kind*)_ω*ness*)_ω +*ish*), but the reverse order is possible (e.g. ((*monk+ish*)_ω +*ness*)_ω).

Raffelsiefen (2024) also goes further on the phonological characteristics of cohering and non-cohering suffixes by noting that non-cohering suffixes have a reduced potential for contrast as they mostly have the form CVC, with a reduced inventory of possible phonemes (e.g. they may contain only high or central vowels), mostly avoiding marked structures. This is opposed to cohering suffixes, which show a greater diversity of possible contrasts and segmental configurations.

One interesting observation that she makes and that may be an issue for other frameworks, is that there are (rare) cases of non-cohering suffixation involving a bound root. Raffelsiefen notes that, when the suffix is productive¹⁰², we find Final-C effects that indicate non-cohering behaviour (74a) that are identical to those found in non-cohering suffixations with free bases (74b).

(74) a.	/əwn.lɪj/	(on) _ω ly	b.	/'ləwn.lɪj/	(lone) _ω ly
	/'gɔ:m.ləs/	(gorm) _ω less		/'hɑ:m.ləs/	(harm) _ω less
	/'ɔjnt.mənt/	(oint) _ω ment		/ə'pɔjnt.mənt/	(appoint) _ω ment

The phonology of the words in (74a) cannot be captured by an account relying solely on OO-Correspondence constraints as there is no free base to license those phonotactic structures. In a stratal model, one would have to assume that they are lexically stored word-level constructs, in which each formative has gone through a stem-level cycle enforcing the phonotactics of the stem-level.¹⁰³ This would entail assuming that the base is a bound stem, which is rarely found in English, but may be found also in certain types of prefixed words (e.g. *flate* in *inflate*, *deflate*).

Raffelsiefen (2005) also discusses some of the processes gathered in Table 18. Let us briefly summarize her analyses on two of those. On the application of æ-tensing before Class 2 suffixes (e.g. *b*[E]*nner* ‘one who bans’, *har*[E]*ssment*), she argues that, like stress neutrality, some cases can be attributed to non-cohesion (“boundary effects”) and some to paradigm uniformity. Thus, the presence of [E] in *harassment* is argued to be an indicator of the prosodic structure ((*harass*)_ω*ment*)_ω (which ensures the heterosyllabicity of the cluster /sm/ and thus the right syllable environment for æ-tensing), while its presence in *banner* indicates paradigm uniformity with the base *ban*, in which æ-tensing regularly applies. She notes that the

¹⁰² Raffelsiefen cites examples of words with non-productive affixes which have undergone prosodic fusion, even though those suffixes are consonant-initial (e.g. *knowledge*, *hatred*).

¹⁰³ Bermúdez-Otero (2018b) assumes that word-level affixes go through a stem-level cycle, even though they escape the prosodic minimality conditions that apply to lexical words, just like function words do.

distribution of [E] in cohering contexts is “more complicated”, as it can be found in inflected words (e.g. *crasser* (comparative), *classes*, *massing*) but not in words such as *classic*, *classify*, *massive* or *passage*, and she suggests that this may be attributable to semantic drift. Likewise, /g/-deletion is found to apply systematically base-finally before non-cohering suffixes (e.g. *fangless* /'fɑŋləs/, *kingly* /'kɪŋli/) and inflectional suffixes (e.g. *singer* /'sɪŋə/, *singing* /'sɪŋɪŋ/) but more ‘unpredictably’ among other cohering suffixes, as shown in (75).

(75) a.	stronger	/'strɔŋgə/	b.	youngish	/'jʊŋɪʃ/
	diphthongal	/dɪf'θɔŋgəl/		hangee	/'hɑ'ŋɪj/
	diphthongize	/'dɪf,θɔŋ,(g)ɑjz/		Pekingese	/'pɪŋki'ŋɪjz/

The difference between the affixes in (75a) and those in (75b) is captured by differences in their affix-specific grammars. In a stratal model, it would be assumed that the former are stem-level formations while the latter are word-level formations. There are independent arguments for treating the comparative as being sometimes stem-level (see Bermúdez-Otero 2012), the adjectival *-al* is traditionally assumed to be a stem-level suffix, and *-ize* is treated as a dual-level affix, and so the optional /g/ in *diphthongize* could be seen as a direct indication of that ambiguity in its stratal affiliation.

In sum, Raffelsiefen’s proposal seeks to tie a number of observations together using the concept of affix cohesion, so that stress neutrality is seen as only **partly arbitrary**. For non-cohering suffixes, she argues for a **representational** solution to capture stress neutrality, while she uses a **procedural** solution (affix-specific cophologies, differing crucially in the ranking of paradigm uniformity constraints) to capture differences in the stress behaviour of different cohering suffixes. While the analysis of some processes (e.g. stress neutrality, processes related to syllable structure) are unified in stratal analyses, they are analysed as arising from two different sources, non-cohesion and paradigm uniformity.¹⁰⁴

3.5.5.2. *Newell*

Newell (2021) also proposes an analysis of suffix classes using the concept of cohesion, i.e. whether or not affixes are within the syllabification and stress domain of their base. However, the framework that she uses is quite different from Raffelsiefen’s, as she works within **Strict CV** and **Rule-Based Phonology**. She notes that all existing accounts use some form of lexical or morphological diacritics affiliating affix classes or individual classes to specific strata or cophologies, and that this is problematic because it does not provide any explanations as to why those affixes behave the way that they do, and it does not make any phonological predictions. As will be seen below, one key difference with Raffelsiefen’s analysis is that she does not assume that all vowel-initial suffixes are cohering, although cohering suffixes tend to

¹⁰⁴ Raffelsiefen’s (p.c.) reply to this is that the phonological effects of non-cohesion and those of paradigm uniformity are of a different nature: “The latter is sensitive to what sort of relation is at stake (often systematically more extensive requirements for identity in inflectional compared to derivational paradigms), more extensive requirements for identity in semantically closely related words compared to less closely related, etc. As for boundary effects stemming from non-cohesive suffixes, these will persist even after the base becomes obsolete (again, I’m referring to cases like *environmēt*, *ointment*, *wistful*, *gormless*, *ruthless*, etc.).”

be vowel-initial and non-cohering suffixes tend to be consonant-initial. She assumes that the pattern to explain is the one that I reproduce in (76).

(76) **The Level 1/Level 2 pattern**

- a. Affixes may have cohering, non-cohering, or variable phonological behaviour.
- b. All affixes behave as cohering when affixed directly to an uncategorized root.
- c. When an affix attaches to an already-categorized (complex) stem, it will display invariable morphophonological behaviour: it will always behave either as a cohering or as a non-cohering affix.

Note that she restricts her discussion of the phonological behaviour of suffixes to stress, syllabification and footing on the grounds that “these processes display regularities that indicate synchronic phonological activity”. However, the only processes that she discusses with regards to syllabification are the simplification of clusters involving a nasal and syllabic consonants, despite there being a rich literature on other effects of syllabification, as was seen in previous sections.

Instead of arbitrary non-phonological diacritics, Newell proposes that some vowel-initial suffixes have an initial floating vowel. This kind of representation is used in Autosegmental Phonology, for example, to account for liaison in French. It consists in assuming that there are elements on the melodic tier that are not linked to any positions on the CV tier. This is shown for the adjectival suffix *-al* in (77a), while examples of non-cohering suffixes are shown in (77b, c). As can be seen here, vowel-initial suffixes may or may not be cohering.

- (77) a.

	C	V	<i>-al</i> (e.g. <i>parental</i>)
ə	l		
- b.

C	V	C	V	<i>-er</i> (e.g. <i>teacher</i>)
	ə	ɪ		
- c.

C	V	<i>-ly</i> (e.g. <i>fully</i>)
l	i	

The basic idea is that the floating vowel will link to an empty V position which, as Newell argues, are very common in English under the assumptions that:

- words that overtly end in a consonant end in an empty nucleus, as assumed in Strict CV;
- diphthongs are vowel-glide sequences, as assumed here.

The only problematic cases are those in which the base ends in /ə/ (e.g. *algebra* /'aldʒəbɪə/), for which Newell proposes a process of hiatus resolution which generates a diphthong, and so an empty V position for the floating vowel to associate with (e.g. *algebraic* /,aldʒə'bɪɛjɪk/).

Newell follows Giegerich (1999) in assuming that the fact that a suffix takes primary stress is not an indicator of cohesion. Thus, she notes the behaviour of suffixes such as *-ology*,

-ography, -ese, -esque or -ee with regards to the simplification of clusters involving a nasal (78a) and syllabic consonants (78b).¹⁰⁵

- (78) a. wom**ɔ**logy, gang**ɛ**se, gang**ɔ**graphy, king**ɛ**e
 b. puzz[ɹ]ólogy, butt[ɹ]ésque, butt[ŋ]ógraphy, bloss[ŋ]ése

The stress-shifting behaviour of those suffixes would be attributable to their size, and not their class (strata, cophonology or underlying representation).

She uses Phase theory, a cyclic model, to capture faithfulness relations among bases and their derivatives. Like most phonological theories, she assumes that roots do not trigger cycles of interpretation, but that each category-determining affix triggers a cycle of interpretation. Unlike other cyclic models such as Lexical Phonology or Stratal Phonology, it is assumed that there is a single phonological grammar. In her analysis, she relies on the classical generative literature (*SPE* and its successors) and assumes that some material is extrametrical at each cycle/phase (after Hayes 1982; Liberman & Prince 1977). She also assumes that the output of a cycle of interpretation is by default not to be modified by further cycles, unless it is “phonologically justified”, so that extrametrical material tends to remain extrametrical. For example, in *governmentless*, it is assumed that there are three cycles, shown in (79).¹⁰⁶

- (79) a. [[govern]_v Ø]_v → (gʌvə)<n>
 b. [[[govern]_v Ø]_v ment]_n → (gʌvə)<n><mənt>
 c. [[[[govern]_v Ø]_v ment]_n less]_a → (gʌvə)<n><mənt><ləs>

Here, the two affixes are non-cohering, and the last consonant (for verbs and non-derived adjectives) or syllable (for nouns and derived adjectives) is extrametrical when found at the right edge of a cycle (extrametricality is represented by the angled brackets). By contrast, the last affix of *governmental* has an initial floating vowel, which associates with the preceding empty V position. This causes a merger of the affix with the previous domain (-ment), as shown in (80a), yielding the output in (80b), which falls out of the fact that, in Strict CV, no domain can begin with V.

- (80) a.

C	V	C	V	C	V	C	V	-	C	V	C	V	C	V	C	V
g	ʌ	v	ə	ɪ	∅	n	∅	-	m	ɛ	n	∅	t	ə	l	∅
- b. (gʌvə)<ɪn>(mɛn)<təl>

In (80b), the output contains two domains of stress and syllabification as the merger does not affect the initial domain (gʌvə)<ɪn>. This is what is assumed to take place in derivatives with a non-cohering suffix followed by a cohering suffix but, in cases where a word contains two consecutive cohering suffixes, it is assumed that “the entire string will always behave as though

¹⁰⁵ The source of the pronunciations reported here is a small survey conducted by Newell as she reports consulting speakers on some of those words, along with her own native speaker intuitions.

¹⁰⁶ This is different from Stratal Phonology, which assumes that there is a single word-level cycle.

monomorphemic with regards to syllabification and footing”. This means that derivations that contain only cohering affixes will constitute a single domain of syllabification and footing, as opposed to the multiple domains in (80). However, in the case of a two-cycle derivation, derivatives containing cohering suffixes may behave differently from one-cycle formations. The difference between one- and two-cycle derivations can be seen in the analysis that Newell proposes for the widely discussed case of *condensation*. She assumes that the variation in the second vowel, which may be /ə/ or /ɛ/, is attributable to “the variable parses in the morpho-syntax, and not to the variable application of vowel reduction or stress-retention”. Thus *cònd/ə/nsátion* would have the structure $[[[condens]_{\downarrow} -ation]_n]$ and be derived in one cycle, while *cònd/ɛ/nsátion* would have the structure $[[[condens]_{\downarrow} -\emptyset]_v -ation]_n]$ and be derived in two cycles, so that it may retain the vowel of *condéense*.¹⁰⁷

In the case of affixes with a variable behaviour, she assumes that they only behave as cohering when attached to uncategorized roots, and as non-cohering elsewhere. For example, *cómparable* ‘similar’ is assumed to be derived from the root $[compare]_{\downarrow}$ and so *-able* behaves as cohering, while *compáritable* ‘able to be compared’ is derived from the verb $[[compare]_{\downarrow} -\emptyset]_v]$, and so *-able* behaves as non-cohering (cp. *compáre*). As noted in the earlier literature, this non-cohering behaviour correlates with compositional semantics and cohering behaviour correlates with non-compositional semantics.

Therefore, Newell’s analysis of the suffix classes of English is purely **representational** and seeks to provide a **non-arbitrary** classification by assuming different underlying representations for suffixes through the presence or absence of an initial floating vowel. The procedural machinery of the cycle does not play a role in the assignment of main stress, as each concatenation of an affix triggers a cycle of interpretation, and it is only the representation of the affix that determines whether or not syllabification and footing will be affected. Thus, it plays the same role as in *SPE*. However, the empirical basis for this analysis is much more restricted than the ones it seeks to supersede, as Newell only analyses stress and two processes related to syllabification. Further explorations will show whether that analysis of English suffix classes can account for all the processes involved.

3.5.6. *Lexical conservatism*

A set of observations of “lexical conservatism” are related to the issue of suffix classes and have been at the centre of debates between proponents of Stratal Phonology and those of OO-Correspondence. The expression “lexical conservatism” has been introduced by Steriade (1997, 1999), and she defines it as “a set of grammatical preferences against the use of forms that are phonologically novel or lack lexical precedents”. Key examples are shown in (81).

- (81) a. *remédiable* ♦ *rémedy* ♦ *remédial*
 b. *párodiable* ♦ *párody* ♦ **paródial*

¹⁰⁷ As was seen in §2.2, vowel reduction in English is highly variable and, as will be seen in §3.7.2.3, and so the potential variable parses assumed by Newell must interact probabilistically with the other determinants of reduction.

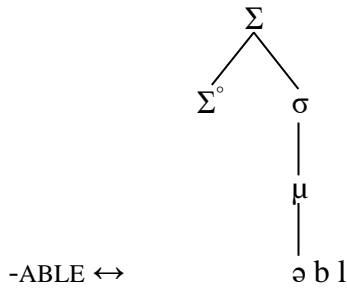
If we compare the stress patterns of *remédiable* and *párodiable*, it can be assumed that the apparent ‘stress shift’ to the second syllable in *remédiable* relative to its morphosyntactic base *remedy* is attributable to the existence of another potential base in the morphological family which is stressed on that syllable, *remédial*. Steriade argues that that word would be selected to improve the stress pattern of the derivative (by reducing the final string of unstressed syllables). Thus, the phonological base may be different from the morphosyntactic base, which is why she calls those kinds of effects “split-base effects”. By contrast, no ‘shift’ happens in *párodiable* because there is no lexical precedent in which the base has second-syllable stress. This has been argued to be a challenge to cyclic containment, i.e. the fact that the phonological base of a derivative should be contained within it, which is a postulate of cyclic theories.

The examples in (81) have often been discussed and reanalysed, notably by Raffelsiefen (2004: 135) and Bermúdez-Otero (2018b), who argue that the proper base for *remédiable* is actually *remédiàte*, because *-able* derivatives often have an *-ate* verb as their base and as they usually truncate the verbal suffix (Kiparsky 2005: 507; Trevian 2015: 235). Therefore, this would not be a proper case of lexical conservatism. Another case that is found in the literature concerns the restrictions on novel *-ize* formations. Raffelsiefen claims that this suffix has a special cophonology which requires the presuffixal syllable to be unstressed. This predicts that *-ize* may not be attached to a base with final stress, unless there is another member of the morphological family which has a stressless syllable in that position. For example, she argues that it is possible to form *Jápanìze* from *Japán* because the base has an allomorph with a trochaic pattern, *Jàpanése*.

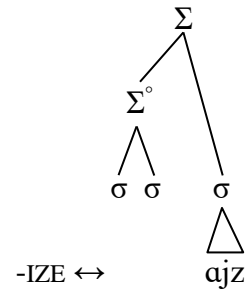
As such cases have been claimed to challenge stratal-cyclic models, in which bases must be strictly contained within derivatives, Bermúdez-Otero (2018b) responds to those criticisms and proposes an analysis in Stratal Phonology in which he assumes that the presence of a stem-level derivative with ‘shifted’ stress in the morphological family of the base “alerts the learner to the availability of the root for stem-level suffixation”. In such cases, derivatives with “dual-level” suffixes such as *-ize* and *-able* are predicted to have stress patterns that are the same as those found in deradical formations. Building on previous literature, Bermúdez-Otero assumes that, in stem-level mode, *-able* is a weak retractor and *-ize* is a strong retractor, which he formalizes as the underlying specifications in (82).¹⁰⁸ This is terminology taken over from Metrical Phonology. **Weak Retraction** means that stress will be placed on the syllable preceding the suffix if it is heavy and two syllables before it if it is light. **Strong Retraction** means that stress will be placed two syllables before the suffix, regardless of the weight of the syllable that precedes the suffix. Here, this is restated with the assumption that a minimal foot (Σ°) must be at least bimoraic. Thus, (82a) implies that *-able* must be preceded by a moraic trochee, i.e. either a stressed heavy syllable or two light syllables, the first one of which is stressed. In turn, (82b) means that *-ize* must be preceded by a disyllabic foot.

¹⁰⁸ This is the way that Bermúdez-Otero uses to capture affix-specific properties: those properties are part of their underlying representations and do not require affix-specific cophonologies, which miss the patterns that they share (see Bermúdez-Otero 2012, 2024).

(82) a. Underlying representation of *-able*



b. Underlying representation of *-ize*



Therefore, *demónstrable* (cp. *démonstràte*, *demónstrative*) is comparable to *deléctable*, and *Jápanìze* is comparable to *récognìze* in that they are all deradical formations. This analysis predicts that there should not be cases in which a derivative with a dual-level suffix has a stress pattern that diverges from that of its local base and that is different from that found in deradical formations. For example, Bermúdez-Otero argues that, although *périodable* has *pèriódic* / ,pr:ɪj'ɔdɪk / in its morphological family, it cannot be stressed *pèriódable*, because that would violate the underlying specifications of *-able*.

The most recent and most extensive investigation of lexical conservatism was conducted by Breiss (2021, 2024). He reports the results of several experiments in which participants are asked to form novel derivatives by adding a suffix to an existing word, which may or may not have a non-local member of the morphological family (called the “remote base”) with a different stress (e.g. *cóntemplàte* + *-able* ♦ *contémplative*; *hábit* + *-ic* ♦ *habítual*), some of the stimuli being taken from a previous experiment reported in Steriade (1997). He finds that stress faithfulness to non-local bases is probabilistic and is only found in English if being faithful to the remote base is “optimizing” (i.e. it reduces the markedness of the stress pattern of the derivative). He also reports that priming the remote base increases the probability of stress being ‘shifted’ in the derivative.

Breiss’s data allows for an empirical evaluation of Bermúdez-Otero’s claim that *-able* will not ‘shift’ stress if it results in an ill-formed foot structure. If we consult his data¹⁰⁹, and focus on *-able* derivatives with a remote base in his Experiment 1 (partly replicating Steriade (1997)) and Experiment 2 (with other stimuli and possible priming of remote bases) in which stress – when shifted – goes onto the pre-suffixal syllable, we get the figures shown in Figure 2.

¹⁰⁹ Canaan Breiss kindly shared his data with me. They should be available online upon publication of Breiss (2024).

THE INTERACTION BETWEEN MORPHOLOGY AND PHONOLOGY

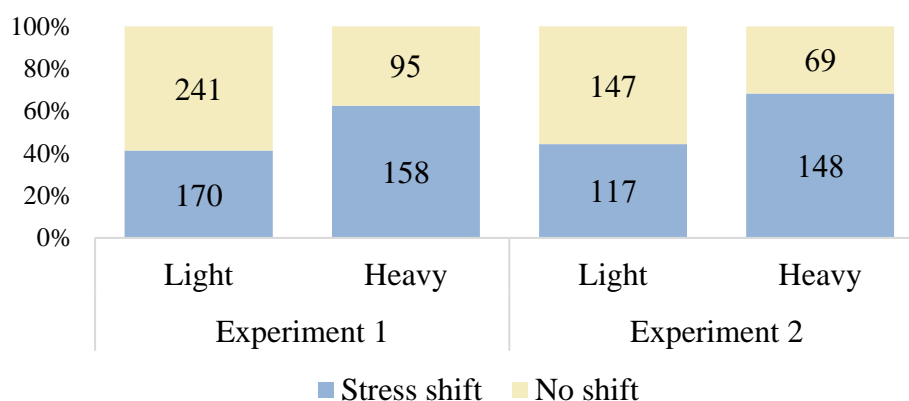


Figure 27. Proportions of ‘stress shift’ in novel *-able* derivatives with a remote base in Breiss’s experiments depending on the weight of the syllable to which stress is shifted

Figure 27 shows that stress is ‘shifted’ more often for derivatives for which the syllable to which the stress ‘shifts’ is heavy (e.g. *ínfiltrâte* + *-able* ♦ *filter*; *párent* + *-able* ♦ *paréntal*), but it is still ‘shifted’ in about 40% of cases in which it is light (e.g. *intúit* + *-able* ♦ *intuítion*; *óffice* + *-able* ♦ *official*). Those results contradict Bermúdez-Otero’s claim and suggest that novel *-able* derivatives do not require *-able* to be preceded by a moraic trochee to ‘shift’ stress.

I will get back to the issue of lexical conservatism in §3.8.3 in the broader context of the discussion of faithfulness relationships beyond the local base.

3.5.7. Summary and perspectives

What can be concluded from this overview of the various analyses that have been proposed to account for the behaviour of different suffix classes? First, the literature shows that the issue goes far **beyond the stress behaviour** of suffixed words, as there are correlates that refer to **syllabification** or **phonetic implementation**. Moreover, even among apparently ‘stress-shifting’ suffixes we find some non-uniformity, as most may affect the foot structure of their base, but some may not (e.g. *-ee* or *-arily*). Therefore, any analysis of suffix classes should take all cases into consideration. The different analyses that have just been reviewed can be summarized as in Table 20 on the two questions in (60).

	Procedural device	Representational device	Behaviour assumed to be arbitrary
<i>SPE</i>	The cycle	Boundaries # +	Yes
Guierrian School (J.-M. Fournier)	Mode of computation / cophonology	None	Yes
Lexical / Stratal Phonology	Stratal affiliation of suffixes The cycle	None	Yes, but partly related to the nature of bases
OO-Correspondence	Output-Output correspondence constraints	None	Yes
Raffelsiefen (2005, 2023)	Output-Output correspondence constraints	Cohesion	Partially, consonant-initial suffixes are neutral
Newell (2021)	The cycle (phases)	Floating vowels	No

Table 20. Summary of analyses of the phonological behaviour of different suffixes

Second, the answers that have been proposed to our questions in (60) on the nature of the difference between types of suffixes vary greatly. We have seen that most analyses assume that there is **at least some arbitrariness** to the difference between stress-shifting and stress-preserving suffixes, and that there are proposals to tie those different behaviours to the phonological form of affixes. Several recent proposals assume that, in the case of affixes with an unstable behaviour, the key lies in the nature of the base (root vs. stem). We have also seen that **purely representational solutions** – the boundaries in *SPE*, the separate prosodification of stress-preserving suffixes – fail to account for the fact that stress-preserving vowel-initial suffixes appear to syllabify with their bases. In *SPE*, stratal models and Newell’s model, the procedural tool of the cycle¹¹⁰ is mainly used in suffixed words to capture Weak Preservation. Other approaches – stratal models and OO-Correspondence – rely on a **procedural** treatment of suffix classes, attributing their different behaviour to their affiliation to different subgrammars (strata or cophonologies), but those fail to capture some of the processes gathered in (73a, b), so that Raffelsiefen argues for an approach which combines representational and procedural solutions. Moreover, a point that the previous sections have not tackled directly is that approaches differ in their treatment of the idiosyncrasies of individual affixes, as the usual

¹¹⁰ For stratal models, that concerns only stratum-internal cyclicity at the stem-level.

strategy in OO-Correspondence is to posit affix-specific cophonologies while Stratal Phonology assumes that the individual requirements of each suffix are part of its underlying representation, and constraints that apply to affixes of the same stratum are captured by the grammar of that stratum (e.g. the avoidance of vowels flanked by two identical consonants at the stem-level, e.g. *optimum* ♦ *optimize* and not **optimumize*).

Therefore, if all the observations that have been reported are correct – and this is something that should be emphasized as many of the processes described in the previous sections rely on a single study (which is sometimes quite old) – then it would appear that both representational and procedural solutions are required to capture the behaviour of different suffixes, possibly along with some lexical storage of affix specificities. Raffelsiefen’s approach seems to meet those requirements, and it would certainly be possible to meet them in Stratal Phonology too. They both employ Prosodic Phonology on the representational side, and the debate surrounding the choice of the procedural mechanism to be preferred goes beyond the issue of suffix classes.¹¹¹ Let us mention how a stratal analysis of the main observations reviewed in the previous sections could function.

First, it would be necessary to assume that the phonotactic restrictions that apply to the right edge of the phonological word (*V]_ω, constraints on final clusters involving a nasal, Final-C effects, syllabic consonants) apply at the stem-level. This can capture the fact that structures that are attested at the end of simplex words can be found before word-level suffixes.¹¹²

(83)	<i>longish</i>	<i>largeness</i>	<i>battling</i>
	[[_{WL} [[_{SL} lɔŋg]] [[_{SL} ɪʃ]]]]	[[_{WL} [[_{SL} lɑ:dʒ]] [[_{SL} nəʃ]]]]	[[_{WL} [[_{SL} batəl]] [[_{SL} ɪŋ]]]]
SL	(lɔŋ) _ω ɪʃ	(lɑ:dʒ) _ω nəʃ	(batl) _ω ɪŋ
WL	(lɔŋɪʃ) _ω	((lɑ:dʒ) _ω nəʃ) _ω	(batlɪŋ) _ω

Second, the allophonic rules of Table 18 and (66) apply at the stem-level when their syllabic conditioning environment is met. Examples are shown in (84).¹¹³

(84)	æ-tensing		Dentalization	
	<i>class-y</i>	<i>class-ic</i>	<i>late-er</i>	<i>matter</i>
SL	[[_{WL} [[_{SL} klæs]] [[_{SL} i]]]]	[[_{SL} klæs-ɪk]]	[[_{WL} [[_{SL} leyt]] [[_{SL} ər]]]]	[[_{SL} mætər]]
WL	(klɛs) _ω i	(klæsɪk) _ω	(leyt) _ω ər	(mæt̚ər) _ω
	(klɛsi) _ω	(klæsɪk) _ω	(leytər) _ω	(mæt̚ər) _ω

Third, word-level suffixes are not all prosodified in the same way, and we may use the concept of cohesion to distinguish them. We can assume that certain suffixes are cohering: they

¹¹¹ Some elements were mentioned in §3.2.

¹¹² The analysis for the simplification of consonant clusters as being a stem-level process is proposed in Bermúdez-Otero (2011) and Bermúdez-Otero & McMahon (2006).

¹¹³ Northern Irish Dentalization is analysed as a stem-level process in Bermúdez-Otero (2011) and Bermúdez-Otero & McMahon (2006).

integrate directly into the phonological word of their base. As the evidence reviewed in previous sections suggests, and following Bermúdez-Otero (2011)¹¹⁴, this is necessary for most vowel-initial suffixes as resyllabification takes place when they attach to consonant-final words. The evidence put forward by Raffelsiefen suggests that consonant-initial are non-cohering and so I will assume that they are attached to a higher phonological word projection, so that the base remains in its minimal projection.¹¹⁵ Then, during phonetic implementation, reference to minimal phonological word projections will be made so that words with a composite structure may behave differently from those with a single phonological word. This is illustrated in (85).

(85)	<i>joyful</i>	<i>rifle</i>	<i>shameful</i>	<i>triumphal</i>
	[[_{WL} [[_{SL} dʒɔj]] [[_{SL} fəl]]]]	[[_{SL} ɪɑj fəl]]	[[_{WL} [[_{SL} ʃɛjm]] [[_{SL} fəl]]]]	[[_{SL} tɹɑjəm f-əl]]
SL	(dʒɔj) _ω - fəl	(ɪɑj fəl) _ω	(ʃɛjm) _ω - fəl	(tɹɑjəm fəl) _ω
WL	((dʒɔj) _ω ° fəl) _ω '	(ɪɑj fəl) _ω	((ʃɛjm) _ω ° fəl) _ω '	(tɹɑjəm fəl) _ω
Phonetic implementation	Longer diphthong Strong [f] Less reduced [ə]	Shorter diphthong Lenited [f] More reduced [ə]	No misalignment of gestures [mf] Less reduced [ə]	Misalignment of gestures [mʰf] More reduced [ə]

Moreover, as there is evidence that certain autostressed suffixes pattern with word-level affixes (see (78)), they will be assumed to be word-level affixes. As to their prosodification, the evidence is quite scarce, but some suggests that they sometimes prosodify with their base, and sometimes not, as shown by the behaviour of voiceless stops before *-ology*, *-ography* or *-ee*, which may or may not be aspirated (Albright 2008). This suggests that they may occasionally form phonological words. Thus, a suffix like *-ee* would sometimes be found at the stem-level when it triggers base allomorphy (e.g. *expellee* / ɛkspe'liː/ ♦ *expel* / ɪk'spel/) or when it attaches to roots (e.g. *amputee* ♦ *amputate*), and sometimes at the word-level, where it may or may not form a separate phonological word (*invitee* (ɪnvajt)_ω(iː)_ω or (ɪnvajtiː)_ω).

Then, let us discuss three particular cases and how they can be accommodated in the present analysis. First, a possible issue is the absence of preservation of base-final unstressed /ɪj/ before word-level suffixes like *-ly* and *-ful* (see fn. 101). It is possible that somehow those affixes are cohering when attached to bases that end in /ɪj/. There are two possible sources for this:

- A process of dissimilation targeting /ɪlj/ sequences and which generates /əlɪj/. In turn, this may lead listeners to interpret words such as *happily* /'hapəlɪj/ as having the structure (*happily*)_ω. However, this explanation can only be invoked for *-ly* and not for *-ful*.
- What Raffelsiefen (2007) calls “High Frequency Fusion”. This is the process by which high-frequency words with a complex prosodic structure fuse into a single phonological word (e.g. *cupboard* (kəp)_ω(boːd)_ω → (kəboːd)_ω). Thus, it is possible that a few words

¹¹⁴ Bermúdez-Otero’s proposal concerns all “stray syllables” though, and no examples of consonant-initial affixes are given.

¹¹⁵ The reason for this different prosodic integration might be, as Raffelsiefen argues, due to constraints on the alignment of prosodic boundaries with morphosyntactic boundaries and syllabification, but it may also be due to the fact that consonant-initial are generally more easily segmentable as they are found in derivatives that are less frequent than their bases, they are quite productive and they often generate word-medial clusters that are not found in simplex words (Hay & Baayen 2003).

containing *-ful* and *-ly* with a base ending in /ij/ have undergone High Frequency Fusion, and that this has then undergone lexical diffusion, so that nowadays language users interpret those suffixes as cohering when attached to bases that end in /ij/.

Diachronic evidence may enlighten us on how to analyse this phenomenon. Second, Newell's claim in (76b) about the "Level 1/Level 2 pattern" must be wrong, as non-cohering consonant-initial suffixes appear to behave that way even when they are attached to bound roots (see the cases in (74a) such as *gormless*). As was suggested in §3.5.5.1, a stratal analysis may have to assume that bound roots such as *gorm*, *ruth* or *grate* are analysed as bound stems triggering a stem-level cycle, and thus behave like the words with free bases in (85).¹¹⁶ Third, the prominence shifts observed in derivatives in *-arily*, *-orily* and \bar{V} *Cable* seen in (63) appear to be different from stem-level 'shifts', as argued by Newell (2021) for *-able*. Those cases involve suffixes that are cohering, as they are either vowel-initial or involve *-ly* after /ij/. Considering that the foot structure of those words does not change with the shift, the change only affects foot labels (i.e. strong/weak), and this may be attributed to the fact that a constraint requiring the head of the phonological word to be aligned with the right edge of the phonological word is getting stronger (see Lindsey 2019: Ch. 23 for other evidence pointing in that direction).

The stress shifts observed in adjectival derivatives in *-atory* appears to be different from the three suffixes above, as shown by diachronic evidence reported by Castanier (2016). He shows that the diachronic evolution of stress in *-atory* adjectives is that shown in (86), which is adapted from Castanier (2016: 543).¹¹⁷

(86) Diachronic shift towards antepenultimate stress in *-atory* adjectives

Phase 1: Four-syllable derivatives

- 1a) Deverbal adjectives (e.g. *migratory*) – from EPD1
- 1b) Deradical adjectives (e.g. *gustatory*) – from EPD14

Phase 2: Derivatives of five syllables and more – from EPD13

- 2a) Primarily derivatives with a five-syllable lapse (e.g. *classificatory*)
- 2b) Secondarily derivatives with a four-syllable lapse (e.g. *adulatory*)

Phase 3: Derivatives expected to resist the change

- 3a) Change involving 'true' stress-shift in derivatives with a disyllabic base with final stress (e.g. *accusatory* /ə'kjʊwzətəɪj/ → /,akjʊw'zɛjtəɪj/ ♦ *accuse* /ə'kjʊwz/) – from EPD14
- 3b) Change involving 'true' stress-shift in *sùpererógatory* → *sùperèrógatory*

¹¹⁶ However, suppletive roots with word-level suffixes (e.g. *better*) are arguably stem-level formations as shown by evidence from Northern Irish Dentalization (e.g. ['bɛɾɹ] 'good.COMP' cp. ['bɛɾ] 'one who bets'; Bermúdez-Otero 2011). Bermúdez-Otero (2012) argues that this analysis can be extended to cases in which the simplification of clusters involving a nasal does not occur before comparative *-er* (e.g. *longer* /'lɒŋgə/, *stronger* /'strɒŋgə/).

¹¹⁷ EPD refers here to different editions of the *Cambridge English Pronouncing Dictionary*.

As can be seen in (86), the first phase of the change involved a shift in deverbal adjectives whose bases are disyllabic. Castanier argues that this was a way for those derivatives to regain stress faithfulness with their base, whose stress had shifted from the first to the second syllable before the XXth century (e.g. *migràte* → *migráte* then *migratory* → *migrátory*). This possible alternative form of the double suffix *-atory* provided a way to preserve the form of *-ate* as /ɛjt/, but also a way to repair the long sequences of post-tonic unstressed syllables in longer derivatives. In those cases, variants in /- 'ɛjtəɲj/ might be seen as word-level formations in which the rightmost foot takes the main stress as in *nècessáryly*. However, the difference is that the older pronunciation in *-atory* adjectives is /-ətəɲj/, and so we could expect an intermediate stage in which there is a post-tonic secondary stress (e.g. *ádulatory* → **ádulàtory* → *àdulátory*). Although Castanier does not discuss such a possibility, Wells (2008) does report such pronunciations. Another element which may have contributed to the emergence of antepenultimate stress in *-atory* adjectives is the existence of a corresponding *-ation* noun.¹¹⁸ However, the further extension of the change to deradical adjectives and in derivatives in which footing differs from that of their base shown in (86) may suggest that some speakers are lexicalizing *-atory* as /- 'ɛjtəɲj/.

Clearly more work is required to understand suffix classes fully, and especially suffixes with an unstable behaviour. This is what we had set out to do with Marjolaine Martin, as the study that we have started focuses on *-able*, *-atory*, *arily*, *-ent*, *-ive* and *-ize*. We want to establish to what extent properties deemed characteristic of the stem-level correlate with one another, and to find out when different diagnostics conflict. We also want to test out Bermúdez-Otero's (2018b) prediction that apparent cases of lexical conservatism involving dual-level suffixes only produce the patterns found when those suffixes are attached to bound roots and only occur when there is a member of the morphological family that has a stress pattern that differs from that of the local base and matches that of the derivative.

3.6. Pretonic secondary stress in stress-shifted derivatives

The literature largely agrees on the fact that pretonic secondary stress placement in stress-shifted derivatives (e.g. *originálicity* ♦ *oríginal*) depends on the position of stresses in that base word (Bermúdez-Otero 2012; Bermúdez-Otero & McMahon 2006; Burzio 1994: 228; Collie 2007, 2008; Guierre 1979: 335; Halle & Kenstowicz 1991; Halle & Vergnaud 1987; Hammond 1989; Kiparsky 1979; Pater 2000; Schane 2007). This kind of preservation has been found to interact with other factors that are known to influence secondary stress placement such as the weight of syllables or the general tendency to avoid stress clashes.

While I was preparing my dissertation, I published a first paper (Dabouis 2016c), in which I studied a preliminary and partial version of the dictionary-based dataset I was going to use for my dissertation, focused on 1,450 suffixal derivatives. The aim of the paper was to

¹¹⁸ Arndt-Lappe & Ben Hedia (2019) report an elicitation study in which participants were asked to produce *-atory* derivatives in which the suffix is preceded by at least two syllables. They show that the frequency of corresponding *-ation* nouns is a predictor of antepenultimate stress in the *-atory* derivative, and so they suggest that several members of the morphological family may be used to compute the stress of *-atory* derivatives (this point will be discussed again in §3.8.2). It would be interesting to evaluate this claim diachronically to establish whether the existence of a (high frequency) *-ation* noun favoured stress shift to *-átory*.

establish to what extent syllable weight (measured in different ways) interacts with Weak Preservation. Even if this is the paper of a young researcher, with no detailed statistical analysis, I believe it contains some interesting results. I never found any effect of weight that would override stress preservation (e.g. *cháracter* ♦ **charàcterizátion*), contrary to what is predicted by the OT constraint hierarchy proposed by Pater (2000). However, I did find a minor effect of syllable weight related to the occurrence of the /021(-)/ pattern (e.g. *depàrtméntal* ♦ *depàrt*), which can be analysed as “exceptional” stress preservation.

I analysed this pattern in more detail in my dissertation, and went even further in a subsequent paper (Dabouis 2019), which focused on 291 suffixal derivatives whose local base has second-syllable primary stress. In that dataset, 32 may be stressed /021(-) (see examples in (87a)) while the remaining 259 may only be stressed /201(-)/ according to the dictionary (see (87b)).

- (87) a. addressee, collectivity, elasticity, departmental, fermentation, reflectivity...
 b. abolition, charismatic, detestation, infectivity, relaxation, unanimity...

This time, the study included a rigorous statistical analysis of the predictors of the /021(-)/ pattern, and it confirmed that its occurrence is limited to derived words.¹¹⁹ The results also show significant effects of the closedness of the first two syllables and the relative frequency of the base and the derivative, such that the /021(-)/ pattern is more likely if:

- The first syllable is open;
- The second syllable is closed;
- The base is more frequent than the derivative.

Moreover, I found that the statistical model improved when the frequency of the most frequent embedded base was taken into account rather than that of the local base. For example, in the case of *connectivity*, the more deeply embedded base *connect* is more frequent than the local base *connective* and so the frequency of *connect* was used in that alternative model. I have explored those aspects of paradigmatic relationships further in subsequent work, and those are discussed in §3.7 along with the general issue of segmentability.

In that paper, I developed a MaxEnt analysis in which those probabilities interact with probabilities in lexical access that a word will be accessed as a whole word or in a decomposed manner (as base + suffix). This was implemented using Hay & Baayen's (2002) “parsing line”, which represents a ratio of base and derivative frequencies above which items are more likely to be accessed through a decomposed route. The details of that analysis are given in §3.7.2.1.

Another phenomenon that I have been interested in regarding stress-shifted derivatives is the regularity of Weak Preservation in contexts in which it should be unconstrained by the constraint on adjacent stressed syllables. The issue is of particular interest as the studies that were available when I started my dissertation reported cases of preservation “failure” (Collie 2007, 2008; Wenzky 2004), i.e. cases in which the derivative may have secondary stress on a syllable that is not (main) stressed in the base (e.g. *antícipate* ♦ *ànticipátion* ~ *anticipátion*). Moreover, Collie found that this is affected by relative frequency so that a more frequent base

¹¹⁹ This observation has been made before by Collie (2007: 79), Hammond (1999: 329), Kager (1989: 171) and Pater (2000) and it presents a particular theoretical interest for Stratal Phonology as it presents a challenge to Chung's Generalization. This will be discussed at the end of §3.7.2.1.

relative to the derivative reduces the chances of stress preservation failure in that derivative. Her analysis was based on dictionary data, but there were several issues with the nature of her dataset, mainly an insufficient control of possible morphological biases and the limited size of the dataset. Thus, additional work on the issue seemed necessary. In my dissertation, I studied stress preservation extensively using dictionary data, which revealed that this is by far the dominant pattern. Stress preservation indeed applies nearly across-the-board, even when there are several stresses to preserve (e.g. *sèntimèntalize* ♦ *sèntimèntalizátion*) or if the base shows variation (e.g. *compútable* ~ *cóputable* ♦ *compùtability* ~ *còmputability*). Strikingly, there are no cases of preservation failure if the stress to be preserved is on the first syllable, and so cases of preservation failure systematically lead to initial secondary stress. This comforts the idea that initial secondary stress is the default in English. In my dissertation, I tested Collie’s hypothesis and failed to confirm her results regarding the role of relative frequency. As for exceptional stress preservation, I expanded the dataset that I had and did some further testing. This new study led to a collaboration with Sabine Arndt-Lappe, who had also been working on the same issue with speech data from an elicitation study. This collaboration led to a joint paper, which will be discussed in §3.7.2.2 as it directly tackles the issue of segmentability.

3.7. The segmentability hypothesis

In several of the studies that I have conducted, I have been interested in what is sometimes called the “segmentability hypothesis”, which predicts different phonetic and phonological behaviours for complex words depending on how strongly their internal structure is perceptible. In this section, I present the initial proposal by Hay (2001, 2003), I detail some of the studies which I have conducted which sought (sometimes among other things) to test the segmentability hypothesis and I conclude with suggestions on how we can go beyond it.

3.7.1. Hay’s model

For some time, the lexicon was viewed as a “collection of the lawless” (Di Sciullo & Williams 1987), which means that it contains only what cannot be computed by rule. This means that complex words are not stored in the lexicon and are computed by language users every time that they are used or encountered. Then, there were debates between theories arguing that complex words are all stored (**full listing**, e.g. Butterworth (1983)) or decomposed (**full parsing**, e.g. Taft & Forster 1975). Now there seems to be a broad consensus that at least some complex words are stored as wholes (see recent overviews in Arndt-Lappe & Ernestus (2020), Aronoff & Sims (2023) and Embick *et al.* (2022)). Thus, more recent models, called “**dual-route**” models, assume that both listing and parsing are available options (see e.g. Baayen & Schreuder 1999; Caramazza *et al.* 1988; Frauenfelder & Schreuder 1992; Schreuder & Baayen 1995; Wurm 1997). While earlier models argued that the absolute frequency of a complex word is what determines its likelihood to be accessed as a whole, Hay (2001, 2003) argues that the

relevant measure to predict which of the two routes is most likely is **relative frequency**.¹²⁰ The assumption behind this proposal is that accessing a word in the lexicon leaves traces in the form of “**resting activation**”, and the more often a word is accessed, the higher its resting activation. In turn, a higher resting activation leads to a quicker access. Thus, when the two routes of lexical access – **direct** and **decompositional** – race against one another, what determines how quick a route is is the resting activation of the constituents involved. I reproduce Hay’s example of the word *insane* in Figure 28.

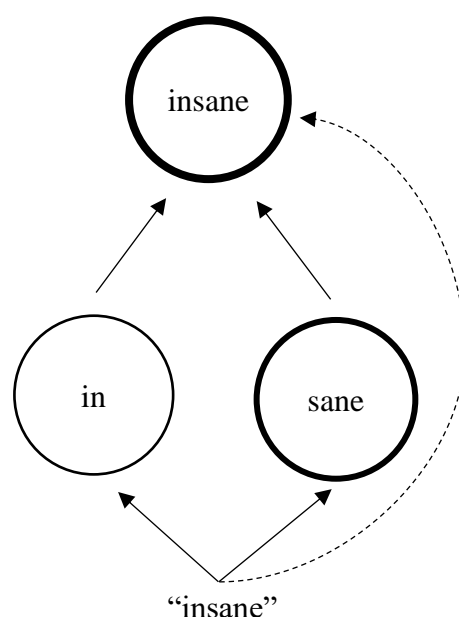


Figure 28. Schematized dual-route model after Hay (2001). The solid lines represent the decomposed route and the dashed line represents the direct route. Resting activation levels are represented by the thickness of the circles (frequencies from the *British National Corpus (BNC)*: *sane* (289), *insane* (360)).

Figure 28 should be read from the bottom and up following either the decomposed route (solid lines) or the direct route (dashed line). We can see that *insane* has a higher frequency than its base *sane*, and so the model predicts that the direct route will be faster than the decomposed route.

I was first acquainted with this hypothesis through Collie's (2007, 2008) work on Weak Preservation (see §3.6) and so I have been interested in testing it in various studies, which are detailed in the following sections. In the literature, several studies have explored the predictions of the segmentability hypothesis in phonetics and phonology (see the overviews in Plag & Ben Hedia (2018) and Stein & Plag (2020)) and effects are not always found, especially when segmentability is measured through relative frequency. The hypothesis is particularly interesting because it predicts when derivatives are likely to drift away from their bases. For example, the higher rate of base-final /t/-elision in *exactly* as compared to *abstractly* may be accounted for by the fact that the former is more frequent than its base (*exactly* (10188) > *exact* (2205)¹²¹) but not the latter (*abstractly* (11) < *abstract* (2148)). Thus, following Hay's

¹²⁰ Note that subsequent works on the segmentability hypothesis have also explored other measures of segmentability such as semantic transparency or affix productivity (Ben Hedia & Plag 2017; Plag & Ben Hedia 2018).

¹²¹ Those figures are token frequencies taken from the *BNC*.

hypothesis, it could be argued that the difference lies in the fact that *exactly* is accessed through the direct route more often than *abstractly*, so that the latter tends to preserve the form of its base more than the former.

3.7.2. Weak Preservation

In §3.6, I started reviewing some of the work that I had done on Weak Preservation. I further discuss this issue by reviewing studies that I have conducted that explored the role of segmentability in Weak Preservation.

3.7.2.1. Exceptional stress preservation

The first one is Dabouis (2019), which focuses on exceptional stress preservation (e.g. *département* ♦ *départméntal*), and which I have already started discussing in §3.6. As was seen in that section, I found an effect of the relative frequency of the base and its derivative. Following Collie (2008), I analysed this using Hay's model of lexical access and assumed that the fact that stress preservation is more common among derivatives whose base is more frequent than them is a sign that those words are accessed through the decomposed route in lexical access. On the other hand, derivatives that are more frequent than their base tend to be accessed through the direct route, and so that could explain why they are more likely to diverge from the stress pattern of their base. In the formal analysis developed in that paper using MaxEnt, this was implemented by assuming that the computation of stress in derivatives that are accessed through the decomposed route has as an input 'free base + suffix' (in which the free base contains its prosodic specification¹²²) as in (89a), while those that are accessed as whole forms have a simple input with its prosodic specification, as in (89b).¹²³ The constraints used in the analysis (slightly adapted from the paper) are shown in (88).

- (88) a. MAX- Σ
The output correspondent of an input foot head must be a foot head.
- b. ALIGN(ω , L; Σ , L)
The left edge of the phonological word must be aligned with the left edge of a foot.
- c. *CLASH
Avoid adjacent stressed syllables.
- d. WEIGHT-TO-STRESS(VC)
Closed syllables must be placed in the position of foot heads.

¹²² This follows from the assumption, adopted in that paper, of Stratal Phonology according to which the outputs of stem-level phonology are stored nonanalytically, i.e. as unanalysed wholes fully specified prosodically (Bermúdez-Otero 2012, 2018b).

¹²³ The weights used for the different constraints here were computed manually as it was not possible to use a software that would have automatically calculated those weights because the model developed here also has to include the probability for a derivative to be accessed through the decomposed or the direct route, as explained below.

The examples in (89) represent derivatives whose bases have stress on their second syllable and have the structure LH(-), in which L represents an open syllable and H a closed one.

(89) a. Decomposed route

Input: /LH́(-)/ free base + suffix	*CLASH 5	MAX-Σ 4	WTS(VC) 3	ALIGN-L 1	H	p(grammar)
/σ(σ̀)(σ̀.../	1			1	-6	0.73
/σ̀(σ̀)(σ̀.../		1	1		-7	0.27

b. Direct route

Input: Listed /LH́σ(-)/ derivative	*CLASH 5	MAX-Σ 4	WTS(VC) 3	ALIGN-L 1	H	p(grammar)
/σ(σ̀)(σ̀.../	1	1		1	-10	0
/σ̀(σ̀)(σ̀.../			1		-3	1

As can be seen in (89), assuming that the input is different for each route of lexical access yields quite different output probabilities for each candidate. Having as input a listed derivative with /LH́σ(-)/ decrease the harmony of the /σ(σ̀)(σ̀.../ candidate so much (because it violates MAX-Σ) compared to the /σ̀(σ̀)(σ̀.../ candidate that the former is predicted not to occur. However, if the input is an /LH́(-)/ free base and a suffix, the probability of having /σ(σ̀)(σ̀.../ significantly increases as it is now the /σ̀(σ̀)(σ̀.../ candidate which violates MAX-Σ.

This way of dealing with frequencies as different segmentability options is not the only option (others will be discussed in the next section), but I believe that this is the first proposal in which this factor interacts with a probabilistic grammar. In the analysis that I proposed in that paper, the probability for a derivative to have one of the two possible patterns was assumed to depend on the probability for that derivative to be accessed through the decomposed route or the direct route and that generated by the MaxEnt grammar, as shown in (90).

$$(90) \quad p(\text{contour}) = (p(\text{grammar}) \times p(\text{input} = \text{free base} + \text{suffix})) + (p(\text{grammar}) \times p(\text{input} = \text{listed derivative}))$$

In this exploratory model, the constraint weights used in (89) were kept constant and the probability for direct or decomposed access was inferred from the data for two subsets of the data: those that have a frequency configuration that is above Hay & Baayen's parsing line (= more segmentable derivatives) and those that are below (= less segmentable ones). Those two subsets were found to have different probabilities for each of the two routes, under the assumption that this is what makes the difference regarding how they are stressed (as suggested by the results of the statistical analysis). Two models were built to compute the interactions in (90) in those two subsets. Table 21 shows how those models were able to capture the distributions found in the data.

	Above parsing line			Below parsing line		
	p(model ₁)	p(data)	Examples	p(model ₂)	p(data)	Examples
#LL(-) /σ̇σ̇σ̇(-)/	0.10	0.08	<i>addressee, diffusivity</i>	0.03	0.04	<i>debauchee, detainee</i>
/σ̇σ̇σ̇(-)/	0.90	0.92	<i>adoration, gazetteer</i>	0.97	0.96	<i>accusation, messianic</i>
#LH(-) /σ̇σ̇σ̇(-)/	0.59	0.56	<i>adoptee, departmental</i>	0.18	0.17	<i>ellipsoidal, reflexivity</i>
/σ̇σ̇σ̇(-)/	0.41	0.44	<i>affectation, domesticity</i>	0.82	0.83	<i>authenticity, molestation</i>
#HL(-) /σ̇σ̇σ̇(-)/	0.01	0.1	<i>consignor, escapee</i>	0.00	0	∅
/σ̇σ̇σ̇(-)/	0.99	0.9	<i>conferee, embarkation</i>	1.00	1	<i>dictatorial, obstetrician</i>
#HH(-) /σ̇σ̇σ̇(-)/	0.10	0	∅	0.03	0.08	<i>encrustation</i>
/σ̇σ̇σ̇(-)/	0.90	1	<i>acceptation, existential</i>	0.97	0.92	<i>dispensation, exultation</i>

Table 21. Probability of the /σ̇σ̇σ̇(-)/ pattern in the models compared to the distributions found in the data

Thus, it was possible to develop a probabilistic model that captures the data quite well, using a rough approximation of how segmentability might function using frequency. Another aspect of segmentability that is shortly explored in the paper is the segmentability of suffixes. Indeed, Hay & Baayen (2002) and Hay (2003: 137) claim that suffixes may vary in their resting activation levels based on how segmentable the derivatives that contain them are. This means that the segmentability of a derivative may not depend just on the base-derivative frequency ratio, but also on the intrinsic segmentability of the affix itself. In the dataset found in the data, it was indeed found that exceptional stress preservation was more likely in words that contain suffixes for which there are more items that are above Hay & Baayen's (2002) parsing line. In other words, the results suggest that exceptional stress preservation is also influenced by the characteristics of individual suffixes, notably their segmentability. This can be illustrated with the two extremes in the data: *-ic* has only 8% of the derivatives that contain it in the dataset that are above the parsing line and no derivative in *-ic* can be stressed /σ̇σ̇σ̇(-)/, but *-ee* has 70% of the derivatives that contain it in the dataset above the parsing line, and 50% of the derivatives in *-ee* can be stressed /σ̇σ̇σ̇(-)/.¹²⁴ This aspect of the segmentability hypothesis has not been studied as extensively as base-derivative frequency ratios or semantic relationships and much remains to be done to investigate the effects of affix segmentability. I will get back to this study when I present the last part of the results that include distant bases, and the model that I proposed to incorporate them into Hay's model.

Let us conclude this section by noting that exceptional stress preservation may be an issue for Stratal Phonology. As was seen in §3.2.2, Chung's Generalization predicts that patterns found in stem-level derivatives should also be found in monomorphemic items (and vice versa). Bermúdez-Otero (2024) gives the examples in (91) to illustrate this using metrical examples.

- (91) a. σ̇σ̇σ̇σ̇σ̇(...) derived in two cycles *orígin* → *orìginálicity*
 derived in one cycle *apòthe-ósis, Epàminóndas*
- b. σ̇σ̇σ̇σ̇σ̇(...) derived in two cycles *comp/ə/nsate* → *comp/ə/nsation*
 derived in one cycle *const/ə/rn-ation, Gorg/ə/nzola*

¹²⁴ This may be further complicated by the fact that *-ee* may be representationally different from the other suffixes in the dataset (*-al*, *-ation*, *-ition*, *-ity*, *-ic*), as was seen in §3.5.

vs. $\sigma\sigma\sigma(\dots)$	derived in two cycles derived in one cycle	<i>condénse</i> → <i>cond/èn/nsation</i> <i>ost/èn/nt-ation</i> , <i>chimp/àn/nzee</i>
c. $\sigma\sigma$	derived in two cycles derived in one cycle	<i>tòrmént</i> _v → <i>tórm/è/nt</i> _N cf. <i>cýpr/ə/ss</i> <i>cónt-òid</i> , <i>wís/è/nt</i>
d. $\sigma\sigma\sigma(\dots)$	derived in two cycles derived in one cycle	<i>régulàte</i> → <i>régulat-òry</i> (Am.) cf. <i>infláme</i> → <i>inflám-atòry</i> <i>véterin-àry</i> (Am.), <i>cátamaràn</i>

Thus, Chung’s Generalization would predict the existence of monomorphemic items with the stress pattern / $\sigma\sigma\sigma(-)$ /. However, no such items are to be found. As argued by Bermúdez-Otero (2024), this is probably not a significant issue for the theory because the process is quite unstable as 28/32 words that may have the / $\sigma\sigma\sigma(-)$ / contour also have a clashless alternative contour / $\sigma\sigma\sigma(-)$ /, and those are already a minority among the words that are candidate for the / $\sigma\sigma\sigma(-)$ / pattern (32/291).

3.7.2.2. Stress preservation failure

In 2017, I exchanged with Sabine Arndt-Lappe (Universität Trier), and it turned out that we had both been working on different kinds of replications of Collie’s (2007, 2008) study of Weak Preservation. Sabine Arndt-Lappe had spoken data from an elicitation study, and I had a dictionary-based study. We decided to give a joint presentation at the 2018 International Morphology Meeting (Arndt-Lappe & Dabouis 2018), which eventually became a joint paper (Arndt-Lappe & Dabouis submitted). The study on dictionary data will be referred to as ‘Study 1’ and the production study as ‘Study 2’.

As I mentioned in §3.6, I had not been able to replicate Collie’s results on relative frequency in my dissertation, with a more controlled dataset. In Study 1, I expanded the dataset beyond the one used in my dissertation to include derivatives from proper names. The dataset contains 277 derivatives with a base that has stress on its second syllable (e.g. *anticipation* ♦ *antícipate*; *majoritarian* ♦ *majórity*; *Napoleonic* ♦ *Napóleon*), 21 of which showed non-systematic stress preservation. Once again, no effects of base frequency were to be found in the binary logistic regression analyses, be it absolute or relative frequency. Inspired by the results of my study of exceptional stress preservation (seen in the previous section), I also tried to include more deeply embedded bases in the frequency computation, and still no effects were to be found. An alternative measure of segmentability, semantic transparency, was tested¹²⁵ but no effects could be found either. All that I could find were effects of the relative weight of the

¹²⁵ It is difficult to find an objective measure of semantic transparency, and so we tried to use a measure that is both non-subjective and replicable. We used a general online dictionary, Dictionary.com (<https://www.dictionary.com>) for that purpose. We coded as ‘opaque’ derivatives for which the base does not appear in the dictionary definition, and as ‘transparent’ the derivatives in which the base does. This may be imperfect but provides an approximation on the transparency of the semantic relationship between the base and derivative. We tried the same measure using the *OED*, but that did not affect the results, as the variable remained non-significant. The other effects remain if that variable is removed from the model altogether.

first two syllables and of prefixation, so that failure to preserve the stress of the base on the second syllable is more likely if:

- The first syllable is at least as heavy as the second (e.g. *ambassadorial* /*ambasə'do:ɪjəl*/; *participation* /*pɑ:tɪsɪ'peɪʃən*/ vs. *appreciation* /*ə,pɹɪjsɪj'ejʃən*/, *respectability* /*ɹɪ,spɛktə'bɪlətɪj*/). This was interpreted following Bermúdez-Otero's (2012) proposal that heavier surrounding syllables may compromise the perception of stress and open the way for preservation failure.
- The first syllable is not a monosyllabic opaque prefix (e.g. *totalitarian*, *antagonistic* vs. *impetuosity*, *evaporation*). This was seen as being consistent with the view that such prefixes are “stress-repellent” (Fudge 1984), and an element of confirmation that those prefixes should be visible to the phonology (see §3.3.1).¹²⁶

Thus, no effects of segmentability could be found in the dictionary data.

Study 2 focuses on spoken data, in which target words were presented in carrier sentences taken from the *Corpus of American Soap Operas* (Davis 2011), as in (92).

- (92) a. Behind all the irrationality, there was a kind of logic in what he did.
 b. The key to success in life is adaptability.

The 56 target words were similar to those used in Study 1, except that the data also contained words with semantically transparent prefixed words (e.g. *implausibility*, *irregularity*, *unsuitability*) and they had only one of the two suffixes found in Collie's study, *-ity* and *-ion*. There were 200 filler items to make sure that participants were unaware of the purpose of the study. The participants were 30 native speakers of British English (17 female, 13 male), aged between 18 and 61 years old, and 27 of which were university students. Secondary stress was rated auditorily by three trained raters who were unaware of the aims of the study, and an observation was kept in the dataset only if at least two raters agreed with one another. This left 1,561 observations.

The first observation is that systematic second-syllable stress is much more marginal (25%) than observed in dictionary studies (74% in Collie (2008) and 92% in Study 1), although second-syllable stress is the majority option for most participants. The factors that were tested in this study were the same as those used in the dictionary data, with additional ones related to segmentability. As certain words were analysable as prefix + free base, we included additional segmentability measures (frequency and semantic transparency – implemented as in Study 1) that made reference to what we called the ‘nominal base’. This was based on the fact that, for those words, two segmentation options are available (e.g. *irrationality*: *irrational* + *-ity* or *ir-* + *rationality*), and we cannot exclude that participants will use both. Moreover, a variable SUFFIX was included to the model to include potential suffix-specificities (see previous section).

¹²⁶ Note that this result is a potential challenge for theories like Stratal Phonology which assume a form of “Bracket Erasure” after the stem-level phonology has applied. Indeed, in that theory, the outputs of stem-level phonology are assumed to be stored nonanalytically, and so the phonology should not have access to morphological information in subsequent cycles. This would predict that there should not be differences between prefixed and non-prefixed derivatives, and yet we do find one.

The data was analysed in two subsets, words without transparent prefixes and prefixed words only (transparent or opaque), using Generalized Linear Mixed Models.

In the subset of the data which excluded transparent prefixes, the presence of an opaque prefix is strongly correlated with the relative weight of the first two syllables and so those variables were tested in different models, in which they were both highly significant predictors of the position of secondary stress. Significant effects of segmentability were found in both models, with semantic transparency being significant in both, and base frequency being significant in the prefix model. Those results are thus in line with Collie's.

In the prefixed subset, which comprises words in which the first syllable is listed as a prefix in the *OED* and in which what the prefixes is attached to is attested as an independent word, there are very interesting segmentability effects, as can be seen in the partial effects plot in Figure 29.

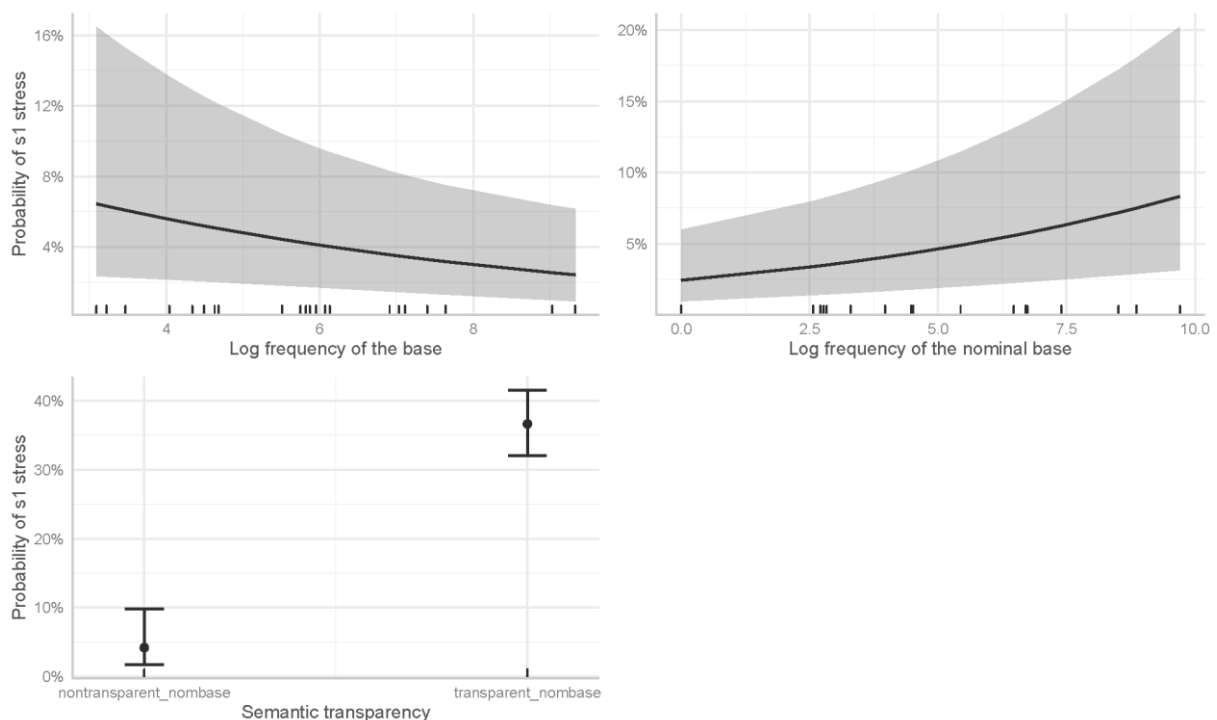


Figure 29. Partial effects plot for the prefixed subset (n=616; other significant predictors include derivative frequency)

Figure 29 shows that there are significant effects of the frequency of both potential bases and that there is an effect of the transparency of the semantic relationship between the derivative and the nominal base, such that first-syllable stress (i.e. preservation failure) is more likely if:

- The (adjectival or verbal) base is less frequent;
- The nominal base is more frequent;
- The semantic relationship between the derivative and the nominal base is transparent (e.g. *im-possibility*, *co-operation* vs. *ac-countability*, *re-liability*).

Once again, the effects of the frequency of the base are consistent with Collie's results, but the effects observed regarding the nominal base refine our knowledge of segmentability as they show that segmentation options that do not fit traditional morphological analyses (which are based on the selectional restrictions of affixes, i.e. the syntactic categories of the bases that they normally attach to) seem to be used by language users.

Thus, while there are clear segmentability effects in Study 2, such effects could not be found in Study 1. In the paper, we propose possible reasons for this discrepancy. The main one has to do with the fact that the words used in the two studies differ considerably. First, there were no words containing semantically transparent prefixes in Study 1 but there were some in Study 2. As was seen above, the existence of an alternative morphological parse appears to make a difference, as nominal base frequency and the transparency of the semantic relationship between the derivative and the nominal base were found to be significant predictors of the position of secondary stress. This can probably be accounted for in terms of informativity: if the nominal base is very frequent, then the prefix is more informative than in derivatives with low-frequency nominal bases. Then, the derivatives used in Study 1 contained a greater variety of suffixes than those used in Study 2. As was seen in the previous section, this may interact with individual base-derivative frequencies. Moreover, the words used in the two studies had quite different frequency profiles, as the derivatives and bases used in Study 1 were overall less frequent than those used in Study 2. If segmentability effects result from lexical storage, then it is to be expected that they will be more difficult to find in a dataset for which the words are less frequent, as low-frequency words are usually not stored.

To conclude on that paper, we found strong effects of segmentability in Study 2, and the absence of such effects in Study 1 should not be taken to mean that segmentability does not affect Weak Preservation. However, Study 2 still has a limited size (only 56 derivatives and 30 participants) and it would be necessary to replicate the study with more derivatives. Such a broader study would need to include more diversity in the stimuli so as to be able to disentangle the effects of relative weight and those of prefixation. It would also be useful to have more items with nominal bases to further confirm the findings reported in the paper. Finally, it might be fruitful to cross those observations from elicitation studies with more naturalistic data from existing corpora or from YouGlish.

3.7.2.3. *Pretonic vowel reduction in stress-shifted derivatives*

In this section, I will detail the part of the study on pretonic vowel reduction which was introduced in §2.2.2 and which focuses on stress-shifted derivatives, which is still taken from Dabouis & Fournier (to appear). As was seen previously, the study includes derivatives whose base has primary stress on the syllable of interest, either initial pretonic (e.g. *vítal* ♦ *vítality*) or intertonic (e.g. *infórm* ♦ *informátion*). Those words are of particular interest because they have been at the core of claims on the cyclic nature of phonological computation since *SPE* (see the discussion of the *condensation* – *compensation* pair in §3.5.1). However, empirical investigations of that process are scarce. One of the only ones is Hammond (2003), who studies *-ation* derivatives in which the second syllable is closed by a sonorant and which have a base with second-syllable stress (e.g. *condémn* ♦ *còndemnátion*), and he finds that vowel reduction in the second syllable is more likely if the base frequency is higher and if the derivative

frequency is higher. However, Hammond's statistical analysis was found to be unreliable by Collie (2007: 182-186). There have been suggestions that what might be relevant is not absolute frequency but relative frequency, so that full vowels are more likely if the base is more frequent than the derivative (Kraska-Szlenk 2007: §8.1.2), in line with Hay's model. Moreover, the discussion of this kind of stress/vowel preservation always focuses on the intertonic syllable of words such as *condensation* and none has looked at cases such as *vitality*. As we have seen, vowel reduction is more common in the intertonic position than in the initial pretonic position, and so it may be interesting to study the initial pretonic syllable too, as such preservation effects might be more easily detectable.

Another aspect of vowel reduction which can be investigated in stress-shifted derivatives is the role of vowel features, as we systematically have access to a 'source' vowel. In non-derived words, we only know whether the vowel is full or reduced and how it is spelled. As was seen in §2.2.2, the four vowel monographs do not behave alike, as <a> and <e> represent reduced vowels more often than <o> and <u>. Therefore, in order to test the role of vowel features, we coded the data using four binary variables encoding the four features [\pm back], [\pm high], [\pm low] and [\pm round] as shown in Table 22, based on Jensen (2022: 64-66).¹²⁷

		BACK	HIGH	LOW	ROUND
DRESS	/ɛ/	no	no	no	no
TRAP	/a/	no	no	yes	no
LOT	/ɔ/	yes	no	yes	yes
STRUT	/ə/	yes	no	no	no
FLEECE	/ij/	no	yes	no	no
GOOSE	/uɪw/	yes	yes	no	yes
THOUGHT/NORTH	/o:/	yes	no	no	yes
PALM/START	/ɑ:/	yes	no	yes	no
NURSE	/ɜ:/	yes	no	no	no
FACE	/ɛj/	no	no	no	no
GOAT	/əw/	yes	no	no	yes
MOUTH	/aw/	no	no	yes	no
PRICE	/ɑj/	yes	no	yes	no
CHOICE	/oj/	yes	no	no	yes
NEAR	/ɪ:/	no	yes	no	no
SQUARE	/ɛ:/	no	no	no	no
CURE	/e:/	yes	yes	no	yes

Table 22. Vowel features based on Jensen (2022)

The subset for stress-shifted derivatives only contains a few cases of digraphs (23 in the initial pretonic position and 5 in the intertonic position) and only 11 words in which the vowel is followed by /sC/ in the intertonic position. Therefore, those items were excluded from statistical analyses. This leaves 567 derivatives for the initial pretonic position and 184 for the intertonic position.

¹²⁷ As Jensen is mainly focused on American English, the vowels /ə:/, /ɪ:/, /ɛ:/ and /e:/ are not present in his analysis, and so we inferred feature specifications from those of other vowels.

A difficulty that one faces when studying base and derivative frequency effects is that the two are often correlated, and so it might be an issue to use both in a regression model. Therefore, the strategy that was adopted was to test the effects of different frequency measures in four different configurations: derivative frequency alone, absolute base and derivative frequency, relative frequency and relative frequency with absolute derivative frequency. Then, we progressively simplified the models to keep only the variables which improve them and performed verifications that the model did not contain harmful collinearity. This was done with the Variance Inflation Factor (VIF; Zuur *et al.* 2010) using the `vif` function of the `car` package. We did not find any cases in which there was potentially harmful collinearity.

Our results show three robust effects which are consistent with observations found in non-derived words:

- vowels in closed syllables (e.g. *tóxic* ♦ *toxícity*) reduce less than those in open syllables (e.g. *pródigy* ♦ *prodígious*). Those followed by /sC/ were found to have an intermediate behaviour in non-derived words, but it is closer to that of vowels in closed syllables in stress-shifted derivatives (see more on this point below);
- vowels found in derivatives with a higher frequency are more likely to be reduced than those found in derivatives with a lower frequency;
- vowels found in an initial pretonic monosyllabic prefix (e.g. *áccess* ♦ *accéssible*) reduce more often than those found in non-prefixed words (e.g. *áctive* ♦ *actívity*), as was seen in Figure 13 in §3.3.1.1.

The variables meant to test the effects of segmentability show inconsistent effects, as we find effects of semantic transparency in both positions (although the effect is only weakly significant in the initial pretonic position), reduced vowels being less common in transparent words than in opaque words.¹²⁸ However, we only found an effect of base frequency in the initial pretonic position, and that effect goes in the opposite direction to that predicted by the segmentability hypothesis: the more frequent the base, the more likely reduction is. Then, we find effects that are inconsistent across the two positions:

- a strong effect of vowel length (opposing long vowels and diphthongs to short vowels) in the initial pretonic position (long vowels being less likely to be reduced in the derivative than short vowels) but not in the intertonic position;
- we find an effect of vowel features, but those differ in the two positions: [+ back] vowels are found to reduce less than [- back] vowels in the initial pretonic position, and [+ high] vowels are found to reduce more than [- high] vowels in the intertonic position. The effect of [± high] may not be reliable as it only contrasts /ɨj/ and /ɰw/ with other vowels, and the dictionary used in the intertonic position tends to use the symbol /u/ for a stressless realization of /ɰw/ more than the one used in the initial pretonic position. Excluding that variable from the regression models did not affect the contribution of the other significant predictors.

Finally, we do not find an effect of the weight of the first syllable in the intertonic position, unlike what was found in non-derived words.

¹²⁸ The measure of semantic transparency that was used was the same as that used in the study of stress preservation failure. See footnote 125.

Although the effects of segmentability are unclear, it is still possible that having a base that has stress on the relevant syllable disfavours reduction. Therefore, we merged the datasets for non-derived words, opaque prefixed words and stress-shifted derivatives and introduced a binary variable DERIVED (YES/NO) to compare them. As effects of vowel features were found in both subsets of the data, we kept in the variable GRAPHEME, which is the only one that can be used in both subsets, and only monographs were kept in. Those merged datasets contain 3,617 words in the initial position and 587 in the intertonic position. In those datasets, we find effects of syllable structure, derivative frequency, grapheme and prefixation that are consistent with our previous results, and we do find a highly significant effect of the new predictor DERIVED in both positions. Therefore, there is indeed a significant difference between non-derived words and stress-shifted derivatives, as can be seen in Figures 30 and 31.

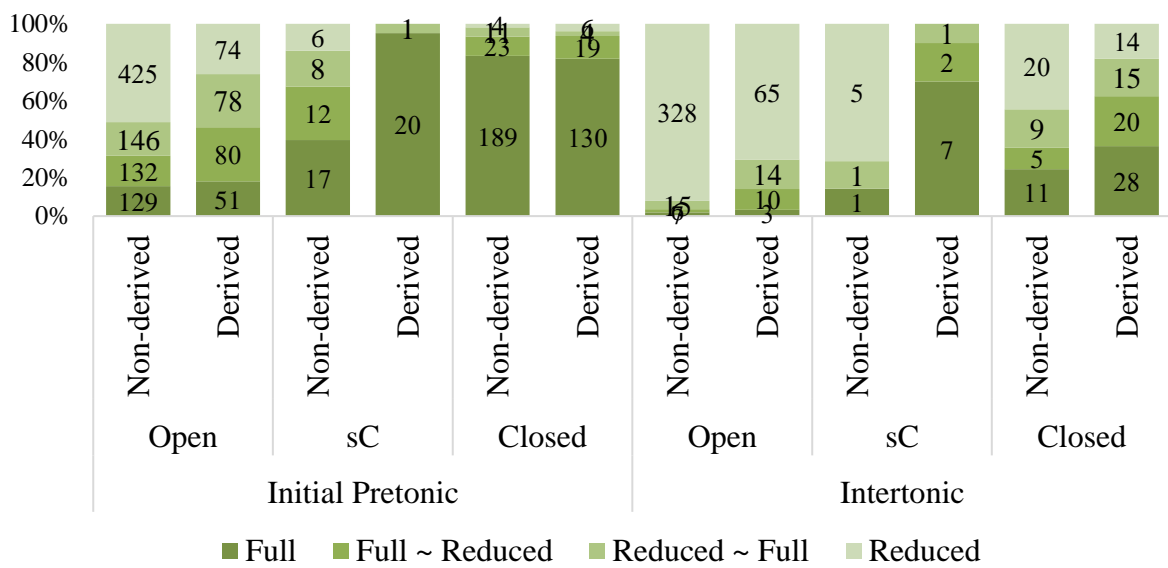


Figure 30. Vowels found for monographs <a, e, o, u> in non-prefixed derived and non-derived words, for both positions and depending on syllable structure

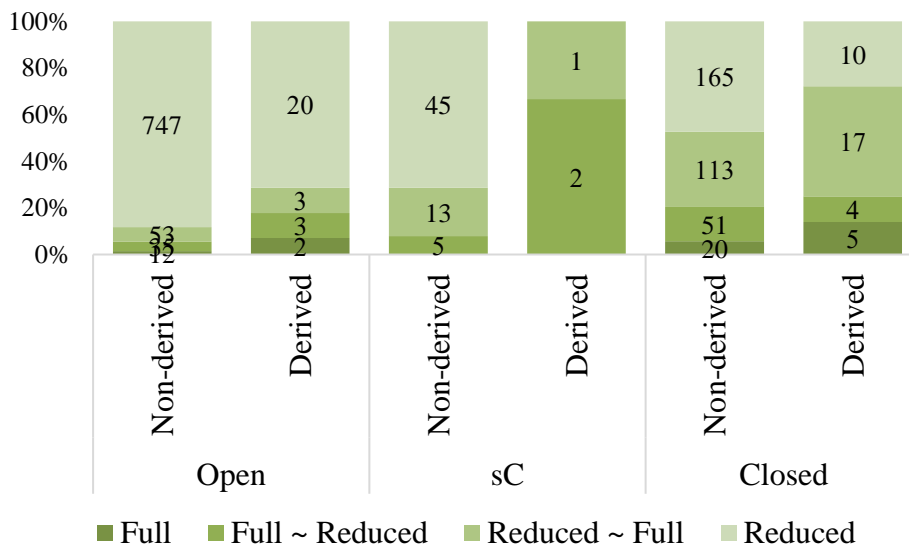


Figure 31. Vowels found for monographs <a, e, o, u> in derived and non-derived words containing an opaque prefix, for both positions and depending on syllable structure

In these two figures, we can see that stress-shifted derivatives systematically have more full vowels than comparable non-derived words. We can also see effects that were described earlier, such as the effect of the position of the vowel, prefixation and syllable structure. The last one shows a peculiar behaviour for vowels before /sC/, which seem to pattern with those found in closed syllables in stress-shifted derivatives while they show an intermediate behaviour in non-derived words. In Dabouis & J.-M. Fournier (to appear), we put forward a possible explanation for this difference. We suggest that /sC/ clusters may show variable syllabification in non-derived words, with heterosyllabic parsing of the cluster favouring preceding full vowels and tautosyllabic parsing favouring reduced vowels. Those two options are shown in (93).

(93) a. heterosyllabic parsing b. tautosyllabic parsing



However, in derived words, two assumptions are required to capture the data. First, /sC/ clusters syllabify heterosyllabically in the base because the syllable that precedes them is stressed, either as a form of coda maximization (Wells 1990) or a requirement that stressed syllables be heavy (Duanmu 2010). Second, the syllabic affiliation of those consonants is taken over from the base, despite other modifications of prosodic structure induced by ‘stress shift’. This idea is similar to Davis’s (2005) analysis of the *capitalistic* vs. *militaristic* pair, in which he assumes that the difference between those two words is due to paradigm uniformity, so that the prosodic structure of the bases *capital* and *military* is preserved into the derivatives.

To conclude, we do find a difference between stress-shifted derivatives and non-derived words, but that difference cannot clearly be connected to segmentability, and it is not a categorical difference. That last point is not really surprising given the fact that vowel reduction is highly variable and influenced by multiple factors. The base frequency effect that we found is not consistent with the segmentability hypothesis and the results reported by Kraska-Szlenk (2007), but they are similar to those reported by Hammond (2003), who finds that a high base frequency favours reduction in the intertonic position. To account for this, he proposes that “the frequency of a complex derived form is a partial function of the frequency of its part”. As will be mentioned in §3.8, it is possible that we need to go beyond the local base and explore frequency effects using frequencies of the whole morphological family.

3.7.3. The gemination of [ɹ]

In 2019, I met with Olivier Glain (Université Jean Monnet de Saint Etienne) to discuss possible joint research projects. We quickly settled on the issue of [ɹ]¹²⁹ gemination, which I had touched in my dissertation (Dabouis 2016b: 505-506). We were quickly joined by Sylvain Navarro

¹²⁹ Throughout this section, I will be using [ɹ] as a ‘cover symbol’ for phonetic realizations of /ɹ/, irrespective of its exact realization.

(Université Paris Cité), whom we contacted because of his expertise on English /ɪ/ (see e.g. Navarro 2017). We decided to use three types of studies, one based on near-spontaneous speech taken from YouGlish (<https://youglish.com/>), an elicitation study and a syllable segmentation task. We piloted the latter two but have only fully explored the first one at this stage. We were helped by a master's student whom I was supervising at the time, Corentin Luc, and for whom I managed to get a small research assistant contract in the first year of his master's degree. He helped us collect and annotate some of the data from YouGlish. We also received valuable help from Paul Lotin, the computer engineer of the LRL, who has allowed us to extract data from YouGlish automatically, and Axel Delarue, the research engineer specialized in statistics. The study has been considerably refined over the course of six presentations in different conferences and workshops (PAC 2021, RFP 2021, LRL seminar 2021, ALOES 2022, “Data SHS” workshop 2022, PAC 2023) to its current state as a submitted paper (Dabouis *et al.* submitted). We have incorporated remarks and suggestions that have been made to us during those presentations, and expanded the dataset because we suspected that some of our initial surprising results were due to a lack of statistical power.

As far as I am aware, I was the first to note that pronunciation dictionaries show that double /ɪ/s at a prefix-base boundary only seem possible in American English, but not in British English. This can be seen in the examples shown in Table 3, which shows pronunciations from two dictionaries that give British and American pronunciations and two American dictionaries.

Dictionary	Entry for <i>irregular</i>	
	British	American
Wells (2008)	ɪ 'reg jʊl ə ,ɪ-, -jəl-	-jəl ə r ,ɪr- ¹³⁰
Upton & Kretzschmar (2017)	(,)ɪ'regjələ(r)	ɪ(r)'regjələ(r)
Kenyon & Knott (1953)	-	ɪ'regjələ-, ɪr'reg-
Merriam-Webster (online)	-	i-'re-gyə-lər ,i(r)-

Table 23. The pronunciation of *irregular* according to four pronunciation dictionaries

Thus, two questions arise:

- Can this difference be confirmed using spoken data?
- Do duration variations for [ɪ] depend on the same variables as those that have been found relevant for other consonants?

There is indeed literature on the issue of gemination, which focuses almost exclusively on nasals. The kind of gemination found in English, which is only found at ‘strong’ morphosyntactic boundaries (e.g. *right time*, *lamp-post*, *unnecessary*), has been called “fake” gemination (Spencer 1996: 25) or “morphological” gemination (Ben Hedia & Plag 2017). Much of the discussion has been around the putative difference between the negative prefixes *in-* and *un-*, the former having been argued not to geminate while the latter having been argued to do so (see Cruttenden 2014: 248; Fudge 1984: 165; J. Harris 1994: 18 and the discussion in Ben Hedia & Plag 2017). This would be due to the fact that *in-* is a Class 1 prefix because it

¹³⁰ Wells (2008) only repeats the part of an entry that have variants. Here, it is significant that he gives /,ɪr-/ as a variant for the first syllable only for American English.

systematically undergoes assimilation, while *un-* is a Class 2 prefix because it does not. However, all the acoustic studies on the issue have shown that this is wrong and that *in-* geminates (Ben Hedia 2019; Ben Hedia & Plag 2017; A. S. Kaye 2005; Oh 2013; Oh & Redford 2012; Raffelsiefen 1999; Videau 2013; Yu 2022), although differences between the two prefixes do exist.¹³¹ Those studies show that nasal duration for those prefixes depends on the productivity of the prefix, on the phonological environment (notably the presence of stress on the following syllable). The effects of the productivity of the prefix have been interpreted as effects of “boundary strength”, which depends on the segmentability of complex words, but no study that I know of has sought to test (relative and absolute) frequency as a predictor of gemination. Our expectation is that [ɪ] duration should be higher in more segmentable derivatives (with higher base frequency and lower derivative frequency), although this could potentially interact with frequency-induced reduction (Bell *et al.* 2009; Clopper & Turnbull 2018; Fidelholtz 1975).

In our study based on YouGlish, we selected 25 words starting with the orthographic string <irr-> varying in frequency, semantic transparency, presence or absence of stress on the second syllable and attestedness of the base. We automatically extracted occurrences of those words from YouGlish and selected only British or American pronunciations. We selected a maximum of 25 occurrences per item per gender and per variety, so a word could have 100 occurrences at the most in our dataset.¹³² The composition of the dataset is shown in Table 4, in which it can be seen that it comprises 19 transparent prefixed words (called ‘derived words’), two of which have several parsing options (*irrevocably*: *irrevocable* + *-ly* or *ir-* + *revocably* and *irresponsibility*: *irresponsible* + *-ity* or *ir-* + *responsibility*), and six opaque prefixed words.

¹³¹ Some studies even show further differences between locative and negative *in-*.

¹³² The details on the constitution of the dataset can be found in Dabouis *et al.* (submitted). The full dataset and R scripts can be found on OSF: https://osf.io/eryca/?view_only=8bd260b0cbc5460f867e2a81ad4ea12b.

THE INTERACTION BETWEEN MORPHOLOGY AND PHONOLOGY

	British English		American English		Total
	Males	Females	Males	Females	
Transparent	269	164	454	345	1,232
irrational	25	16	25	25	91
irreconcilable	20	3	25	25	73
irredeemable	5		25	10	40
irrefutable	8		25	18	51
irregular	25	21	25	25	96
irrelevance	7	1	22	10	40
irrelevant	22	24	24	24	94
irreligious	5		25	6	36
irreplaceable	13	5	25	25	68
irrepressible	7	4	25	15	51
irresistible	24	17	25	25	91
irrespective	25	25	25	25	100
irresponsibility	4		25	7	36
irresponsible	24	15	25	22	86
irretrievable	3	2	17	12	34
irreverence	1	2	16	7	26
irreverent	17	11	25	25	78
irreversible	25	17	25	25	92
irrevocably	9	1	25	14	49
Opaque	51	16	149	98	314
irradiate	3	1	25	3	32
irradiated	5	2	25	8	40
irradiation	4	1	25	12	42
irrigation	12	1	24	25	62
irritate	18	4	25	25	72
irritation	9	7	25	25	66
Total	320	180	603	443	1,546

Table 24. Composition of the dataset in Dabouis *et al.* (submitted)

For each token, we measured manually the duration of [ɪ], [ɪ̥] and the whole word using Praat (Boersma & Weenik 2024). All measurements were checked by another one of the three authors. Figure 32 shows an example of segmentation for a perceptually geminated [ɪ̥] for the first occurrence of *irrelevant* among American male speakers.

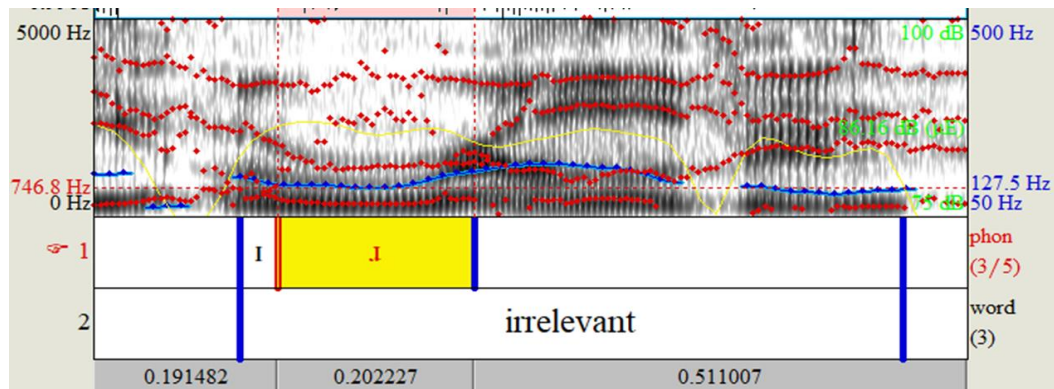


Figure 32. Example of duration measurement for *irrelevant* (USM1)

The data was annotated for the different variables of interest (presence of stress on the second syllable, semantic transparency, gender, variety of English) while controlling for speech rate, so as to evaluate the effects of all predictors in a multiple linear regression model. For frequency variables, which can either be taken as a predictor of the segmentability of derived words or as a predictor of greater reduction in high frequency derivatives, we tested a number of different configurations to see which ones were the best fit for the data:

- Absolute frequencies (base and derivative), either one at a time or both in the same model. As is common practice, they were all log-transformed as $\log_e(x+1)$;
- Relative frequency ($\text{LogFq-Derivative} / \text{LogFq-Base}$), either on its own or along with the absolute frequency of the derivative.

For the models reported below, the lemma frequencies were collected from the *COCAE*. For reasons that will be developed in more detail in the next section, we also tested base frequency measures that were calculated as the cumulated frequencies of all the bases that are embedded in the derivative. For example, for the derivative *irreconcilable*, the closest base is *reconcilable*, which itself contains *reconcile*. When we wanted to test for the potential cumulative effects of base frequencies, the frequencies of both bases were added up to compute base frequency. In the remainder of this section, I report the main results regarding the absolute duration of [ɪ], although the paper also deals with that of [ɪ] and with the relative duration of [ɪ] and [ɪ], and the implications of the study.

The analyses were conducted on the data from each variety of English separately, but the results were largely identical, except for an absence of an effect of word frequency in British English. For clarity of presentation, we thus chose to collapse the two subsets into one ($n=1,546$) and tested for an interaction between word frequency and variety of English, but none was found. We used the same modelling strategy as that described of the study on vowel reduction in §2.2.2.2, and found strong effects for all predictors. Those effects can be seen in the effects plot shown in Figure 33, which represents the effect of each predictor while controlling for other predictors. Note that the 95% confidence interval is represented by the light blue areas for continuous variables (here frequency and speech rate).

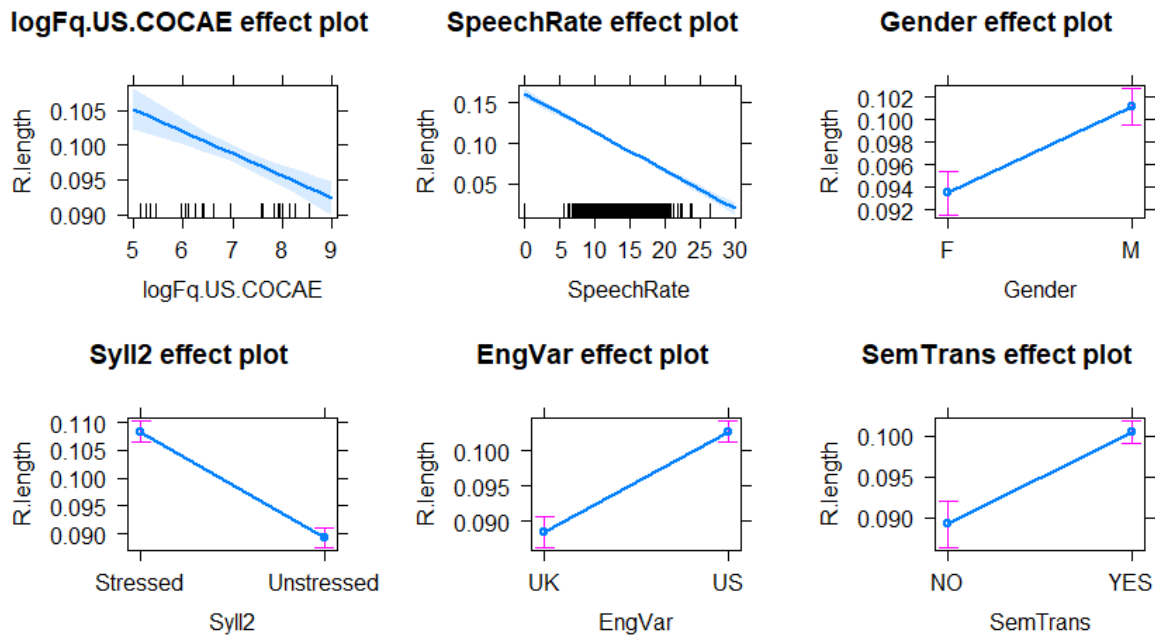


Figure 33. Effects plot for the linear regression model on [ɪ] duration in the full dataset (taken from Dabouis *et al.* submitted)

Figure 33 shows that [ɪ] is longer if:

- the speech rate is lower;
- the frequency of the word is lower;
- the word is semantically transparent;
- the following syllable is stressed;
- the speaker is a male;
- the speaker speaks American English.

Thus, our results confirm the results reported literature on the role of frequency and stress on the following syllable on consonant duration. The effect of semantic transparency is consistent with the segmentability hypothesis, as transparent items are more segmentable, and we do get longer [ɪ]s in those words than in less transparent ones. The effect of gender is new and unexpected, and I will get back to it below. Finally, there is indeed an effect of the variety of English, with longer [ɪ]s in American English than in British English.

Then, we analysed the duration of [ɪ] in the subset of derived words, excluding the two whose present more than one parsing option (n=1,145). This subset is particularly interesting if we want to study more gradient effects of segmentability, measured through frequency. As was mentioned above, different frequency measures were tested in a number of regression models in which the covariates were always the same (and always contributed to improving the model). Those different models were compared on the basis of their AIC, and the model with the lowest AIC is the one which includes the frequency of the derivative and a measure of relative frequency in which the frequency of the base taken to compute the ratio is the cumulated frequency of all embedded bases. In that model, there is also an interaction between derivative frequency and the variety of English. The effects found for the different predictors are shown in Figure 34.

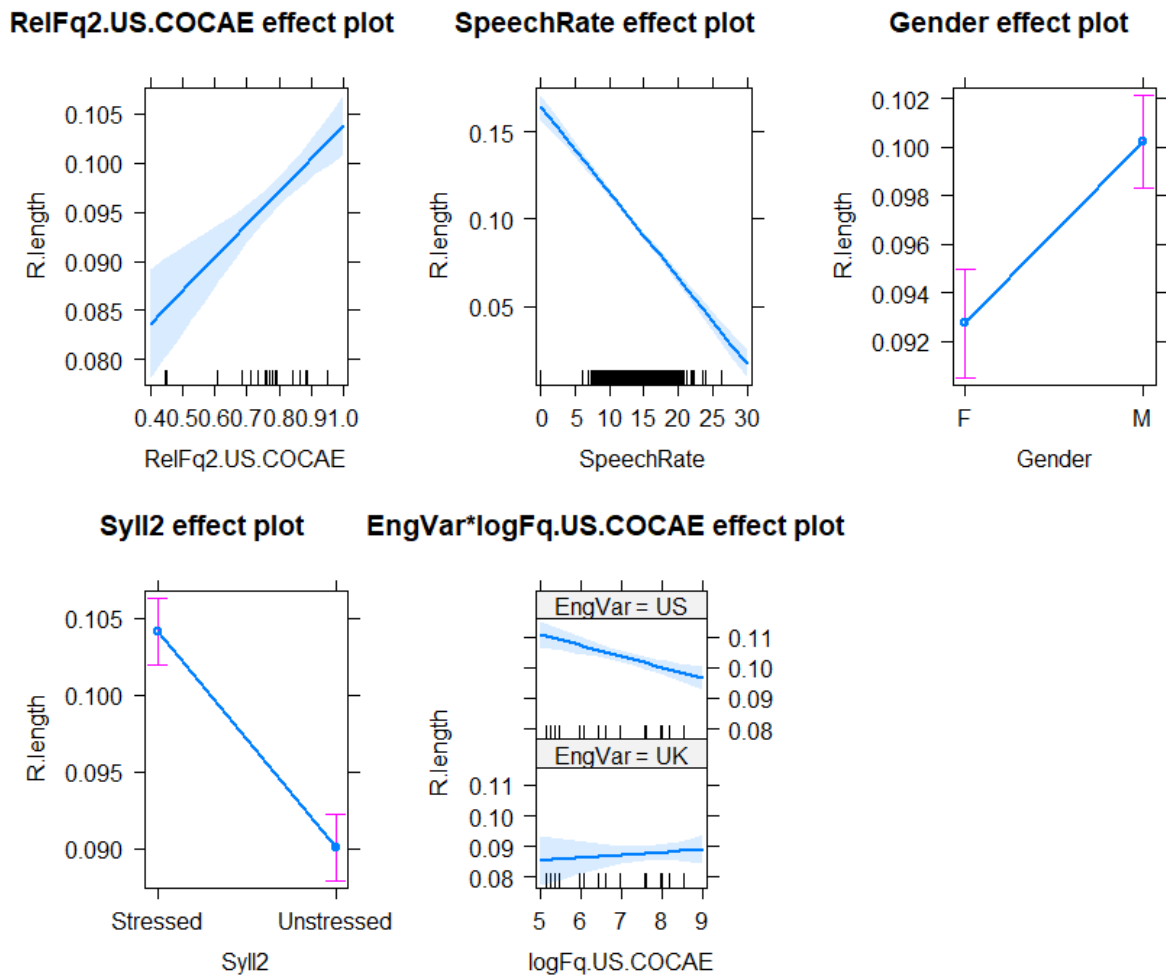


Figure 34. Effects plot for the linear regression model on [ɪ] duration in the derived words (taken from Dabouis *et al.* submitted)

Figure 34 shows that the effects of speech rate, gender, variety of English and presence of stress on the second syllable are the same as those observed in the full dataset. However, we only find a (small) effect of derivative frequency in American English. Most interestingly, the strong effect of relative frequency is as follows: the more frequent the base is relative to the derivative (= the lower the relative frequency), the shorter the [ɪ]. This result was found in all the models in which base frequency (however it was measured) was included. However, because of the small number of types (n=17), it should probably still be taken with caution.

That being said, this result contradicts the predictions of the segmentability hypothesis, as we would predict [ɪ] to be longer with a lower relative frequency, not shorter. Our earlier results partly showed this effect, and we were increasingly puzzled as it was confirmed at every step in the refinement of the study – puzzled because we did not have an idea as to why our results were entirely the opposite of what we were expecting. Then, I came upon a possible explanation when I came across Engemann & Plag’s (2021) study of the duration of plural forms in New Zealand English. In this study, they study the duration of the bare stem (here ‘stem’ refers to the part of the word-form that excludes the inflectional suffix) in plural word-forms such as *boys*, *jars* or *trees* and evaluate the effect (among other predictors) of the absolute frequency of the base and derivative and of relative frequency. They find that the duration of

the bare stem is shorter if the frequency of the plural word-form is higher and if the frequency of the bare stem is higher. They interpret this in terms of paradigm uniformity and frequency-induced reduction: a frequent bare stem tends to be shorter than a less frequent one, this shorter duration is stored in long-term memory, and so it can be found to affect the duration of the plural word-form. In turn, the same effect of frequency on reduction can directly affect the duration of the plural word-form itself. Our results suggest that this kind of paradigm uniformity extends to more distant bases as they also appear to contribute to the reduction of the duration of the [ɪ]. However, Engemann & Plag also report effects of relative frequency that are consistent with the segmentability hypothesis: if the bare stem is more frequent than the plural word-form, then the duration of the stem in the plural word-form is found to be longer. It is possible that the difference is due to differences in the frequency configurations found in our data as compared to theirs. Therefore, our study should be replicated on other consonants, for which more types are available and so a closer investigation of frequency effects should be possible.

One of our aims was to evaluate whether the difference between British English and American English reported in pronunciation dictionaries holds when using speech data. Our results do not allow for a clear-cut answer, but they do show that differences exist. First, we do find that the realizations of /ɪ/ tend to be longer in American English than in British English. However, we cannot take this to mean that gemination is possible in the former but not in the latter, and we do find clearly geminated examples in the British data, such as the one shown in Figure 35.

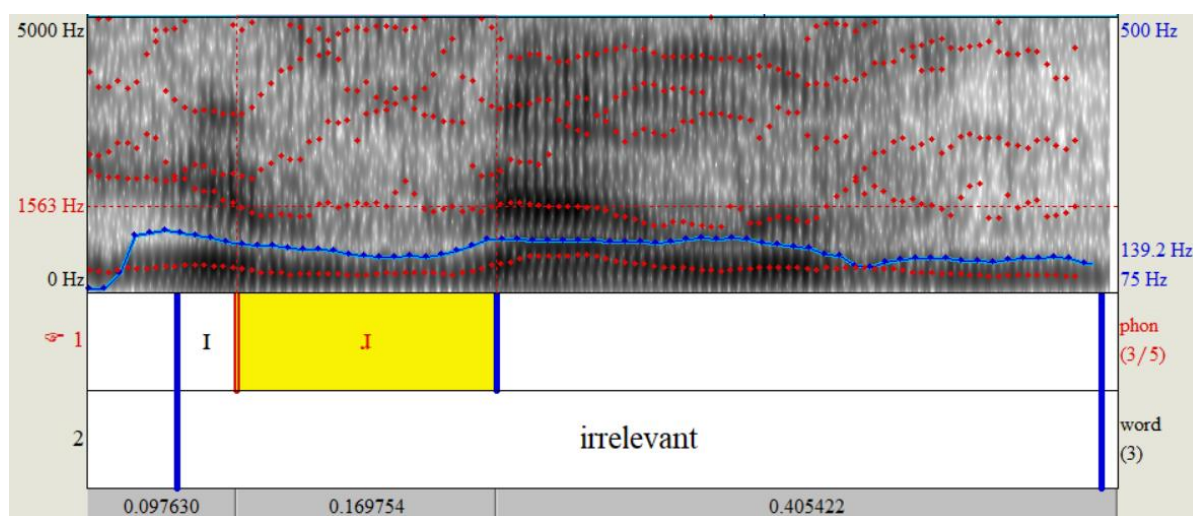


Figure 35. Example of geminated [ɪ] in British English (*irrelevant* - UKM23)

Another difference that we found between the two varieties is that [ɪ] duration is more affected by word frequency in American English than in British English. While we do not have any good explanation for this at this point, we suggest that this could have to do with the differences in the allophones of /ɪ/ and how they are used in different varieties of English (see King & Ferragne 2020), and that techniques such as ultrasound tongue imaging may be helpful in exploring that hypothesis.

Finally, let me conclude with the effect of gender, which was unexpected. In the paper, we suggest a possible explanation (put forward by Olivier Glain) for the observation that male speakers produce longer [ɪ]s than female speakers in our data. It has often been shown in

sociolinguistic research that, “in changes from below, women are most often the innovators” (Labov 1990: 215) and oratory and speechmaking (a type of speech that resembles the data that we are studying, which is often scripted speech and/ or public speech) have been going through a stylistic change in recent decades, so that speechmakers aim at sounding more informal (McWhorter 2003). Thus, it is possible that women are leading this change too, so that they drift away from more formal forms more than men. In our case, longer [ɹ]s could be seen as more formal realizations, under the influence of the orthographic sequence <rr>, and so shorter [ɹ]s may be seen as less formal.

As for possible perspectives for this study, it would be interesting to replicate it for other consonants to ascertain whether the frequency effects can be replicated if we study a dataset with more types (and so more variability in frequency distributions), which would be possible for nasals. It may also be informative to conduct production studies on [ɹ] gemination that are similar to those conducted by Oh (2013) and Oh & Redford (2012), in which speech style was varied. Finally, factors that were not considered here may be tested in future studies on [ɹ] or other consonants, such as the position in the sentence or the presence of emphasis on the prefix.

3.8. Beyond locality and containment

In the last thirty years or so, there has been a debate among phonologists regarding the issue of paradigmatic dependencies and the identification of the members of the morphological family which may act as phonological bases for a derivative. I have worked on those issues in several studies and am currently writing a chapter about them with Sabine Arndt-Lappe in a collective volume about theoretical phonology and the phonology of English edited by Patrick Honeybone and Martin Krämer. There are two key points of controversy: locality and containment. In models which have a cyclic architecture, phonological computation is usually assumed to be **strictly local**, i.e. the only base that is relevant for the computation of a complex word is the local base. However, as will be seen below, models such as Stratal Phonology allow for distant bases (i.e. more deeply embedded bases, e.g. for *connectivity*, this would be *connect*) to be used as bases. I have worked on this issue in the context of my studies of Weak Preservation and will discuss my proposals in §3.8.1. In some cases, among the bases that are contained in a derivative with multiple affixes, **several parsing options** are available and may affect the phonological form of the derivative (e.g. *irrationality*: *ir-* + *rationality* or *irrational* + *-ity*). This issue will be discussed in §3.8.2. Then, there are models that assume that bases need not be contained within the derivative, so that **all members of the morphological family are potential bases**. This raises the issue of choosing between different bases if only one can be used, or that of the importance given to each possible base if several can be used at the same time. Those proposals will be presented in §3.8.3.

3.8.1. Distant bases

In her study of Weak Preservation, Collie (2007: 288-289) mentions that it is possible to have what she calls “leap-frogging” preservation, i.e. faithfulness of the derivative to a distant base. For example, Bermúdez-Otero (2007a) reports that a colleague from Manchester has the

influence of the distant base (if they share the phonological property under investigation) or a conflict between these two influences (if they do not share that property).

2. $FqDistantBase < FqLocalBase$: the local base is accessed directly and then the derivative is accessed through the decomposed route. In this configuration, we do not expect to see a difference between derivatives with only a local base and derivatives with a distant base.

B. $FqLocalBase < FqDerivative$

1. $FqDistantBase > FqDerivative$: the derivative is accessed through the decomposed route but the local base is skipped. This is the configuration in which we would expect leap-frogging preservation.
2. $FqDistantBase < FqDerivative$: the derivative is accessed directly. In this configuration, we expect preservation phenomena to be more likely to fail.

These four configurations therefore correspond to four possible routes of lexical access depending on the access route which is used between each pair of related words, which are schematized in Figure 36. This figure represents the different routes of lexical access described in (95), which are identified as A1, A2, B1 and B2. The figure should be read from the bottom and up following one of the four routes.

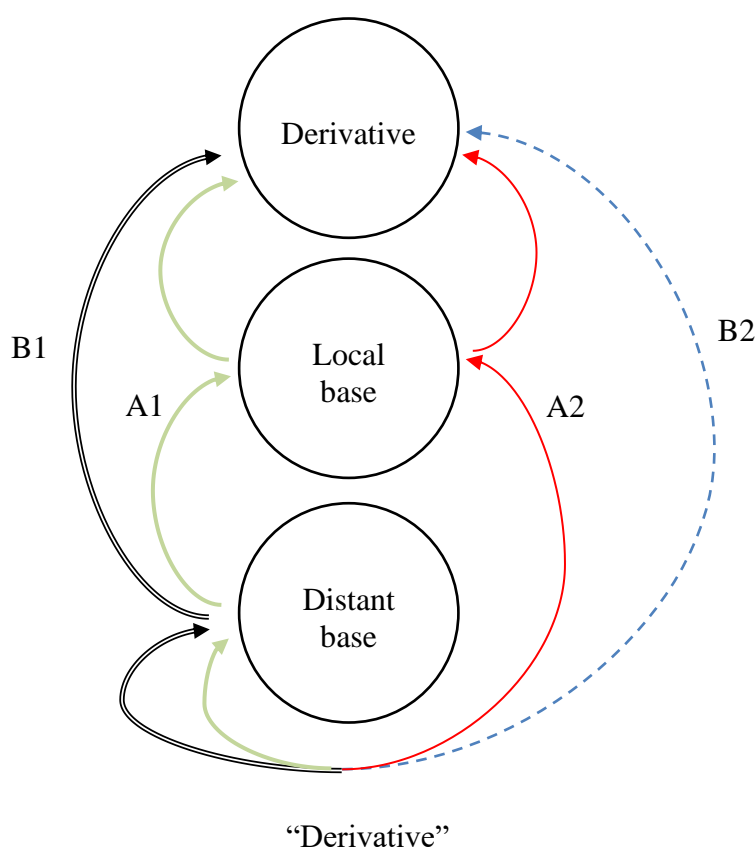


Figure 36. The four possible routes of lexical access with a distant base

Therefore, this model formalizes how distant bases might play a role in the lexical access of complex words, and makes predictions on how that might affect their phonology. Even though it may seem considerably more complex than Hay’s model, I now believe it to be oversimplistic. I am increasingly convinced that morphological relationships between bases and derivatives are gradient and so that segmentability itself is a gradient property of derivatives. The model presented here basically imports gradient morphology into the phonology categorically in the form of a binary choice between two bases. In §3.8.3, I will mention how it may be possible to directly capture the gradience of morphological relationships in phonological computation.

We have also explored the role of distant bases in the paper on Weak Preservation conducted with Sabine Arndt-Lappe which was presented in §3.7.2.2. As a reminder, it deals with derivatives which all have a local base with second-syllable stress (e.g. *imagine* ♦ *imàginátion*). In both studies reported in that paper, we coded whether there was a distant base and how it is stressed. No effects could be found in Study 1 (dictionary data), but a small effect was found in Study 2 (speech data) in the subset that excludes semantically transparent prefixed words. In both models, we find a significant effect of the variable encoding the presence or absence and stress pattern of a distant base. The effect observed in the model in which the relative weight of the first two syllables is used (and not opaque prefixation – remember that those two variables could not be included in the same model) can be seen in Figure 37.

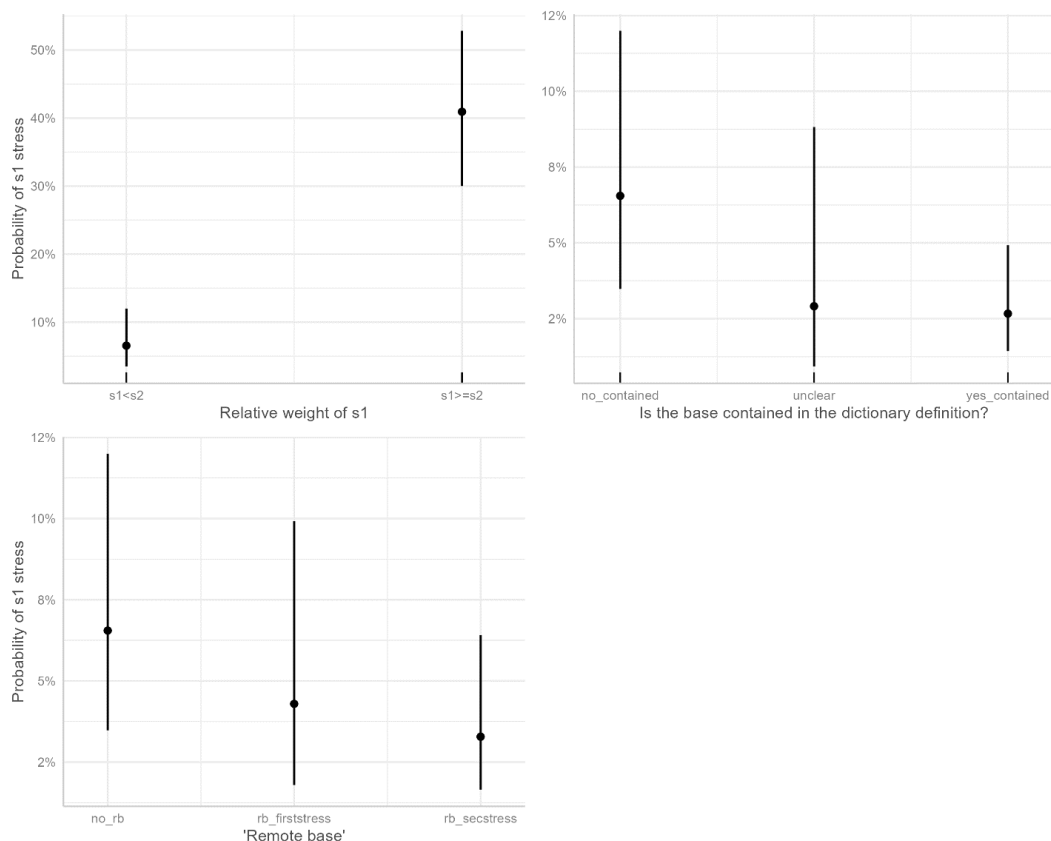


Figure 37. Partial effects plot for the subset without transparent prefixes (n=1,087)

As can be seen in Figure 37, the presence of a distant base with second-syllable stress (e.g. *sustáin* ♦ *sustáinable* ♦ *sustainability*) weakly favours second-syllable stress as compared to

derivatives with a distant base with first-syllable stress (e.g. *súbstance* ♦ *substántiate* ♦ *substantiátion*) or derivatives with no distant base (e.g. **suscept* ♦ *suscéptible* ♦ *susceptibility*). The only significant effect is that having a distant base with second-syllable stress increases slightly the probability of having second-syllable stress in the derivative. Strikingly, derivatives with no distant bases have significantly more first-syllable stress than derivatives that have a distant base with first-syllable stress. This is quite unexpected and should be investigated further in future research. Note that the study was not designed to test this particular aspect and so there are not enough items per category to really be able to see the potential effects of distant bases with some confidence.

In §3.7.3, we also saw that the best model to capture the duration of [ɪ] in prefixed words in *ir-* was the one for which the measure of base frequency that was used was the cumulated frequencies of all embedded bases.

Therefore, those three studies provide some evidence that distant bases play a role in Weak Preservation and call for further studies on the extent to which distant derivatives may be faithful to them.

3.8.2. *Multiple parsing options*

As was seen in §3.7.2.2, in certain cases there are several parsing options for a given derivative, and so several potential local bases (e.g. *impossibility*: *im-* + *possibility* or *impossible* + *-ity*). In that section, we saw some evidence from Weak Preservation that those different parsing options are used by language users as we found frequency effects from both local bases.

Similar effects have also been reported by Arndt-Lappe & Ben Hedia (2019), who study the position of primary stress in *-atory* adjectives in an elicitation study, in which participants were asked to read such words in frame sentences. Those adjectives often have two potential bases: a verbal base and a nominal base in *-ation* (e.g. *oscillatory* ♦ *oscillate* ♦ *oscillation*, *explanatory* ♦ *explain* ♦ *explanation*). The verbal bases are usually stressed on one of their first two syllables while the nominal base always has primary stress on *-átion*. Arndt-Lappe & Ben Hedia tested the effects of the frequency of both bases as predictors of the position of primary stress chosen by participants for *-atory* adjectives, and they found effects of both: a higher base frequency entails a higher probability for the *-atory* adjective to be faithful to that base. This suggests that *-atory* are either derived through the juxtaposition of *-(at)ory* to a verbal base or through affix substitution (*-ion* → *-ory*) to nouns in *-ation* and that, once again, language users may use both parsing options. However, this result already breaches the limits of containment as the *-ation* base is not really contained within the *-atory* derivative, which opens the question of faithfulness relationships throughout the morphological family.

3.8.3. *Faithfulness throughout the morphological family*

As was seen in previous sections, a central point of debate between OO-Correspondence and cyclic models is whether or not there can be **faithfulness beyond containment**, i.e. whether a derivative can be faithful to members of its morphological family which are not the local base or the distant base, i.e. co-derivatives. When we discussed lexical conservatism in §3.5.6, we saw that there is some evidence that derivatives may adopt the stress patterns of members of

their co-derivatives in certain conditions. If we go beyond those two theories, there are frameworks which assume that multiple words that are related to a given derivative in some way may be used to perform computation (e.g. Kuperman *et al.* (2007), and analogy-based frameworks; see e.g. Arndt-Lappe (2015) for an overview). In this section, we will address the issue of the factors that affect the choice of a phonological base and the issue of faithfulness to several bases.

First, let us address the issue of the **factors affecting base selection** or, if we assume that multiple bases may be available simultaneously, the importance of a given member of the morphological family in the computation of a derivative. The three main proposals are the factors shown in (96).

- (96) a. **Resting activation:** Several researchers have proposed that frequency affects the choice of the phonological base (Breiss 2021, 2024; Collie 2007; Dabouis 2019), and frequency is assumed to affect resting activation and thus how available (easily retrievable) words are. Breiss (2021, 2024) also argues that resting activation may be manipulated experimentally through priming and provides experimental evidence that participants stress novel derivatives like primed non-local bases more often than unprimed non-local bases.
- b. **Markedness:** Researchers that have worked on lexical conservatism in OO-Correspondence frameworks often claim that markedness drives base selection (see Breiss 2021, 2024; Stanton & Steriade 2021), so that non-local bases may be selected if being faithful to them improves the markedness of the derivative as opposed to being faithful to the local base (Breiss (2024) calls such remote bases “helpful”). Thus markedness may favour the selection of a local base. However, in cases in which being faithful to a non-local base leads to the derivative being more marked as compared to being faithful to the local base (Breiss (2024) calls such remote bases “harmful”), then markedness may prevent faithfulness to a non-local base.
- c. **Priority to the local base:** We have already exposed the fact that cyclic models only consider bases that are contained within the derivative as possible phonological bases. Even in non-cyclic models, it is usually assumed that the local base has an advantage over other bases. Steriade & Stanton (2020) posit a violable constraint C-CONTAINMENT which assigns violations “to any derivative D whose base is not exponent of an immediate constituent of D”. This is a way to penalize non-local bases and to formalize a form of preference for the local base. Likewise, Breiss (2021, 2024) posits different faithfulness constraints for the local and the non-local base in a MaxEnt model, each with its own weight. When he fits the model to his experimental data, he finds the constraint requiring faithfulness to the local base to have a greater weight than the one requiring faithfulness to the non-local base.

All the analyses that I know of have used several of those factors in interaction with one another in OT models, and it is possible that certain phenomena may only emerge in specific configurations of those different factors. For example, as noted by Breiss (2024), lexical conservatism requires “a sufficiently strong motivating markedness constraint to encourage

deviation from the Local Base, [...] a Remote Base in the paradigm of the Derivative, and [...] a sufficiently weak faithfulness constraint indexed to the Local Base”.

While it is quite straightforward why the first two factors in (96) play a role, one might wonder why the local base would benefit from an advantage. One possible answer is that the local base is usually the one that has the greater degree of formal and semantic overlap with the derivative, and that this stronger proximity is what gives it an advantage, should other bases exist. This suggests that this advantage of the local base actually captures effects of semantic and formal similarity. In my talk at the 2023 MFM (Dabouis 2023b), I explored another possible answer as I looked at a set of 789 stress-shifted derivatives taken from Dabouis (2019) and Dabouis & J.-M. Fournier (to appear). I found that, for half of them, the only other member of their morphological family is the local base, as can be seen in Figure 38.

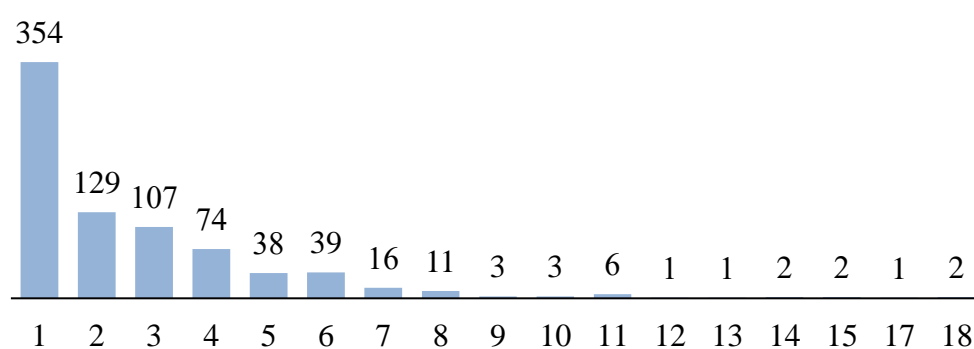


Figure 38. Number of words in morphological family per derivative in Dabouis (2023b)

This can be taken to suggest that there is an advantage for the local base because it is often the only available base, so that language users may adopt a general ‘local base only’ strategy.

In cases in which the possibility for a derivative to be faithful to a non-local base is allowed, the factors described in (96) have been used to decide which other member of the morphological family will be used as the phonological base. The idea that underlies such models seems to be that there can be **only one base ‘at a time’**. One option is that there is only one base in the input, and the choice of which base is determined outside of the phonology (as I have proposed in Dabouis 2019). Another option is that there can be several bases in the input, and that output candidates (in an OT analysis) are evaluated depending on their faithfulness to a single base. An illustration of the second option can be found in Steriade & Stanton’s (2020) analysis of lexical conservatism. They use the constraint C-CONTAINMENT mentioned above, the constraint BD IDSTRESS which assigns violations to syllables which do not have the same stress value in the base and in its counterpart in the derivative and *LAPSE_{LAT}, which penalizes sequences of more than one unstressed syllable in Latinate words. They use the word *gelatinoid* to illustrate how their analysis functions. This word has a local base (B^L), *gelatin*, and a non-local base which they call the remote base (B^R), *gelatinous*. Their tableau for this derivative is reproduced in (97).

(97)	B ^L : g��latin B ^R : gel��tin-ous	100 010-0	BD IdSTRESS	*LAPSE _{LAT}	C-CONTAINMENT
☞ a.	gel��tin ^R -��id	0102			*
b.	g��latin ^L -��id	1002		*	
c.	gel��tin ^L -��id	0102	*!*		

As can be seen in (97), each output candidate may stand in correspondence with one of the input bases, as represented by the superscripts ^L and ^R. The local base is only favoured by C-CONTAINMENT, which is ranked lower than *LAPSE_{LAT}. Thus, *gel  tin  id* is eliminated if it stands in correspondence with *g  latin* (97c), but not if it does with *gel  tinous* (97a). As for candidate (97b), *g  latin  id*, it is eliminated because it violates *LAPSE_{LAT}, even if it does not violate C-CONTAINMENT.

Breiss (2021, 2024) makes a different proposal: output candidates are evaluated in correspondence with **several bases at the same time**. More precisely, the input forms are the different stem¹³⁴ allomorphs found in different bases. For example, in his experiment he has the novel derivative *laborable*, which has a local base *l  bor* and a remote base *lab  rious*. Those two bases correspond to the two stem allomorphs /'l  jb  (ɪ)/ and /l  'b  :ɪ-/. The evaluation proceeds similarly to (97), with the notable difference that correspondence with the stem allomorphs in the input is evaluated through faithfulness constraints that refer to each allomorph (ID-[STRESS]-LOCAL and ID-[STRESS]-REMOTE). Thus, there is no separate constraint penalizing non-contained bases, and so the advantage of the local base is achieved through a greater weight attributed to the constraint requiring faithfulness to the allomorph of the local base. In case the remote base was primed, it was found that the weight of the faithfulness constraint requiring stress faithfulness to the allomorph of that base was higher.

In my 2023 MFM presentation (Dabouis 2023b), I explored the idea that the whole morphological family could influence the phonological form of a derivative. In my 2019 paper, I had seen that taking into account the frequency of the most frequent base improved the regression models as compared to models that included the frequency of the local base. Taking this idea further, I made the hypotheses in (98).

¹³⁴ Breiss does not define this term, but it seems to be used in the same way as *Raffelsiefen*: that which remains once one or more affixes are removed.

- (98) a. If the morphological family of a derivative is homogenous with regards to phonological property \mathcal{P} (e.g. having stress on a certain syllable) and if its members are more frequent than that derivative, then the likelihood for that derivative to have \mathcal{P} is higher.
- b. If the morphological family of a derivative is heterogenous with regards to \mathcal{P} and if its members are less frequent than the derivative, then the likelihood for that derivative to have \mathcal{P} is lower.¹³⁵

In that presentation, I only explored (98a) by comparing models that include the frequency of the local base, models that include the cumulated frequencies of all embedded bases and models that include the cumulated frequencies of all the members of the morphological family that have the same stress pattern. The data that I used are the datasets used in my 2019 paper (§3.7.2.1) and the one used in the study of vowel reduction in stress-shifted derivatives (§3.7.2.3). In the cases in which several bases were used, the cumulated frequencies of bases was calculated as $\log_e(Fq_1 + Fq_2 + \dots + Fq_n + 1)$.¹³⁶ The different models were compared using their AIC, and I found that models were systematically better if non-local bases were taken into account, but that the models which included the cumulated frequencies of all the members of the morphological family did not always perform significantly better than those which included only the cumulated frequencies of embedded bases. However, when a difference was found, it was always the model with the frequencies of the whole morphological family that performed better. In the conclusion of that presentation, I suggested that if such effects are confirmed, they could be formalized using Gradient Harmonic Grammar (Smolensky & Goldrick 2016), in which phonological representations may be gradient, with different levels of “activity”. Here, it could be assumed that activity is related to the resting activation of the different bases (or stem allomorphs), and that violations of faithfulness to an input form are a product of the weight of the faithfulness constraint and the activity of that input form. Such a model would also need a way to account for the advantage of the local base which, as I have suggested above, may have to do with the higher semantic and formal proximity between the derivative and the local base as compared to non-local bases. This means that this advantage of the local base is also gradient and may only be calculated on an individual basis. The details of such an analysis require a considerable amount of work and will need to be explored further in future research.

To conclude, we have seen that there is some evidence suggesting that faithfulness goes beyond locality and containment. The evidence often comes from highly variable processes, and it systematically shows that the local base has a clear advantage over other members of the morphological family. However, the evidence suggests that paradigmatic dependencies may extend beyond the local base in certain conditions. This directly challenges models that restrict paradigmatic dependencies to the local base or to embedded bases. If such evidence continues

¹³⁵ As noted above, faithfulness interacts with markedness so that the outcome for a derivative to have – or to adopt – \mathcal{P} depends on whether \mathcal{P} goes against markedness constraints or not. If \mathcal{P} leads to a more marked pattern, then the question is to know whether the derivative will adopt (or maintain) the marked pattern, i.e. whether faithfulness to members of the morphological family is strong enough to overcome markedness constraints.

¹³⁶ It was suggested to us by Julien Eychenne that it would be better to use the sum of log-frequencies, i.e. $\log_e(Fq_1 + 1) + \log_e(Fq_2 + 1) + \dots + \log_e(Fq_n + 1)$.

to be found, then those models will have to be adapted to account for such dependencies. However, evidence of paradigmatic dependencies involving non-contained bases may not necessarily challenge the whole architecture of a theory such as Stratal Phonology. A crucial aspect of that theory is the assumption that the outputs of the stem-level phonology are stored nonanalytically in the lexicon, and so it may be possible to assume that relationships that go beyond containment are formed between such stored forms, at the stem-level. As the outputs of the word-level are not stored or are stored analytically, we do not expect paradigmatic dependencies to violate containment in word-level forms, and so interstratal cyclicity need not be challenged. However, it may be necessary to get rid of stratum-internal cyclicity at the stem-level and instead assume a more gradient approach to morphological relationships at the stem-level that could accommodate for a broader range of paradigmatic dependencies. What is currently captured through stratum-internal cyclicity may simply be a reflection of the advantage of the local base, and so getting rid of containment at the stem-level may broaden the empirical content of the theory.

CHAPTER 4

Orthography and phonology

4.1. The previous literature and first explorations

Although there is a rich literature in psycholinguistics and reading acquisition on the relationship between orthography and phonology, the topic has been largely ignored from the literature in phonology. We do find *SPE*'s (1968) remark that “English orthography, despite its often cited inconsistencies, comes remarkably close to being an optimal orthographic system for English” and some relatively isolated works that have tackled the issue directly (Giegerich 1992, 1999: Ch. 5; Montgomery 2001, 2005) and the recent extension of the BiPhon model (which will be discussed in §4.4). Giegerich and Montgomery identify the limits of phonological derivation and argue that, for some processes at least, it is necessary for the phonology to have access to orthographic information. Apart from those, mentions of orthographic information in research on English phonology are scarce and only deal with specific processes. Relevant examples are Raffelsiefen (2010: 197), who argues for an orthographic treatment of Velar Softening, or the processes surrounding orthographic consonant geminates, which have been widely discussed (see for example *SPE*: 148; Burzio 1994: 56; Collie 2007: 134; Giegerich 1999: 164; Stockwell & Minkova 2001: 173).

Some have put forward hypotheses as to why there has been such a disinterest. For example, Derwing (1992) and Giegerich (1992, 1999: 155) propose that early linguists such as Saussure and Bloomfield were not interested in orthography because they assumed that “spoken forms alone constitute the object [of linguistic study]” and that others after them have argued that all languages have a spoken form but not all have a written form, so that we should focus on spoken forms. This is consistent with Guierre's (2000) suggestion that that disinterest comes from what he calls a “trend” of the quest for linguistic universals. However, one interested in language universals may certainly want to focus on spoken forms because not all languages have a written form but, as Giegerich (1999: 155) points out, that “tells us nothing about the relationship between spoken and written form in languages that have both”. Moreover, as will be discussed below, there are reasons to believe that languages which do have an orthographic system might behave differently from those which do not.

Within the Guierrian tradition, the reference to orthography has been present since Guierre's first published works and goes largely unquestioned. I have argued, along with others, that the inclusion of orthography is one of the defining features of the Guierrian School (Dabouis 2016b: §2.1; Dabouis *et al.* 2023). In his thesis, Guierre (1979) argues that there are four possible grammars, the grammar of the reader, the grammar of the writer, the grammar of

the listener and the grammar of the speaker, and that the grammar of the reader is the most efficient to formulate generalizations regarding stress placement and the realization of vowels. Within this tradition, the study of spelling-to-sound regularities has been considerably developed by Deschamps (1994) for vowels and Trevian (1993) for consonants. J.-M. Fournier (1992, 1994, 2010) then formalised the whole system in the form of a diagram which relies on the hierarchy of factors which determines the realization of vowels. The impact of research in the Guierrian School has been considerable in the French educational system, as the rules that have been developed in that framework have constituted the basis for the exercises given at the agrégation and, more recently, at the Capes.¹³⁷

The stark contrast that I observed between those different approaches made me think that it was an issue worth exploring, and so it was on this topic that I gave a talk in 2013 at the University of York on an issue that was not directly related to the research I had been conducting for my master's thesis. The proceedings of that conference ended up being my first publication (Dabouis 2014). Although this paper now reads as the first hesitant steps of a young researcher, it covers some of the processes that a model of English phonology that includes orthographic information should absolutely integrate, and it introduces an idea that I still deem relevant.

There is a considerable number of morphologically related words that show vowel alternations of the kind illustrated in (99).

(99)	divine	/dɪ'vɑj̃n/	◆	divinity	/dɪ'vɪnətɪj/
	isobar	/'ɑjsəwba:/	◆	isobaric	/'ɑjsəw'baɪk/
	barbarian	/'bɑ: 'bɛ:ɪjən/	◆	barbaric	/'bɑ: 'baɪk/

There is considerable difficulty in deriving those vowels from one another because, as Green (2007) points out, there is “the problem of determining what the underlying phonemes are whose allophones are [aɪ, ɪ], [i, ε], [e, a], [o, ɑ], because the surface allophones are featurally often more distinct from each other than they are from allophones of a different phoneme”. Thus, several approaches have been proposed in the literature, some assuming a simple connection between vowels that alternate with one another, some deriving vowels from other vowels or more abstract approaches in which different surface vowels are derived from a common underlying vowel (see e.g. *SPE*; Jensen 2022). In my 2014 paper, I propose that those alternating vowels are best seen as part of a series (e.g. /{ɛj, a, ɑ:, ε:}/, /{ɪj, ε, ə:, ɪ:}/). In the Guierrian approach, the term “quality” has been used to refer to groups of vowels sharing a grapheme, and so what connects those vowels is not simply the fact that they alternate, but that they are all written with the same grapheme.

Shortly after, I continued studying this issue by looking at the psycholinguistic literature on literacy and its impacts on the phonological system, looking at the scarce works in the phonological literature that directly address the issue of the relationship between phonology and orthography, establishing the processes that seem difficult to account for without referring to orthographic information and detailing the impacts that integrating orthographic information has on how we can analyse the stress system in English. Those investigations were the topic of two talks in 2015 at the PAC and RFP conferences and were continually expanded to become the

¹³⁷ Those are two competitive exams opening the way to teaching in secondary education.

chapter that I published in the collective volume following PAC 2015 published in 2023 (Dabouis 2023a). In the next section, I review the main arguments for considering that, at least for English phonology, orthography is difficult to ignore and then, in the following section, I detail the processes for which a purely phonological analysis seems difficult to maintain, and for which it is thus necessary to refer to orthographic information. Those two sections will mainly draw on my 2023 chapter. Subsequently, I will turn to the place of orthography in the grammatical system and will finish with studies which I have conducted on pretonic vowels with secondary stress and how I tried to extend some of the generalizations that I found to other environments.

4.2. Orthography is not a side issue

While it is true that not all languages have orthographic systems and that there may be language universals regulating the phonology of all languages, it seems difficult to assume that languages in which a large proportion of the population is literate function in the same way as languages which do not. Likewise, at the individual level, it seems difficult to analyse literate language users as if they had no knowledge of orthography, as if that knowledge had no impact on their phonological system. However, this seems to be a common approach, the most striking being *SPE*: Bermúdez-Otero (p.c.) describes the kind of speaker-listener whose grammar is developed in *SPE* as “hyper-literate”, although the authors of *SPE* do not acknowledge that fact.

This position seems untenable given that the acquisition of literacy is a long and tedious process and is thus a **powerful force to restructure phonological knowledge** (Jaeger 1986; Laks 2005). Moreover, psycholinguistic evidence shows that there is indeed such a restructuring and that this reorganisation of the cognitive system is permanent (Bürki *et al.* 2012; Ehri & Wilce 1980; Perre & Ziegler 2008; Taft & Hambly 1985; Wang & Derwing 1986). Orthographic effects are also well documented in loanword adaptation (e.g. Hamann & Colombo 2017; Neuman 2009: §5) and in second language acquisition (see Bassetti 2023; Bassetti *et al.* 2015 for overviews).

Thus, while I might agree that an orthographically-informed phonology is not ‘pure’ phonology, I contend that integrating orthographic information to study the phonology of English language users comes down to studying an object that is by far more common than studying the phonology of illiterate language users of English. Finally, although orthography is learned later than the bases of the phonology, it does not seem to be less important or salient than phonological information. Bürki *et al.*’s (2012) experiment is enlightening in that regard. In that experiment, adult native speakers of French had to learn noncewords which may be taken to be the reduced variants of words that have lost a schwa. The stimuli were presented auditorily twenty-five times overall several days and once visually, half with an orthographic <e>, half without. Then, they were asked to produce the words orally, and they were found to be more likely to produce a schwa if a word has been presented with an <e>. Finally, they were presented with spoken forms and asked whether they were part of the words that they had learned. The results showed that they were more likely to say yes to a variant with schwa when the spelling that was presented for that word contained an <e>. The authors thus conclude that “a single exposure to spelling following extensive phonological learning is sufficient to induce such a change in previously established phonological representations”.

4.3. The limits of phonology

One of the rare generative phonologists to have argued for the integration of orthographic information in the phonological analysis of English has argued that certain alternations “breach the limits of what can be predicted on phonological grounds alone” (Giegerich 1999: 5). In my 2023 chapter on the issue, I review such alternations and suggest possible ways that orthographic information could be imported in the phonology. In the rest of this section, I briefly discuss some of those alternations.

4.3.1. ‘Unnatural’ alternations

Alternations such as the ones in (99) have been strongly debated in the generative literature because of the difficulty to relate the vowels that alternate with one another. If such alternations are to be captured by the phonology (although some reject this; see e.g. J. Kaye 1995; Szigetvári 2018), we have the following options:

- **A common underlier:** The two vowels derive from a common underlier, which would imply strong modifications to derive the correct surface vowels.¹³⁸
- **Derive a vowel from the other:** One type of vowel is considered basic, and the other type is derived from it. There too, computation bears the heavy load of deriving vowels that are quite different from one another.
- **Surface connections:** The two vowels are arbitrarily marked as alternating with one another, despite their surface dissimilarity.

The third option has the advantage of doing away with excessively complicated derivations or highly abstract underliers, but it raises the question of how the vowels are to be connected. One could assume that simply noting the alternations in words such as (99) is enough for the language user to construct associations between vowels. However, such alternations are not that common, and they occur mainly in educated vocabulary that is probably mainly manipulated by literate language users. Thus, those users have access to an additional cue that connects alternating vowels: they have the same orthographic representation (e.g. both /a/ and /ɛj/ are spelled <a>) and the association between those vowels and their spelling extends far beyond those alternating pairs (e.g. /a/ - <a> is found in non-alternating words such as *cat*, *action*, *fanatic*). There is evidence supporting that option, as the rule of ‘Vowel Shift’ that was posited in *SPE* to derive such alternations was found to be:

- marginally productive;
- most productive for literate language users;
- productive only for alternations which are represented with the same spelling (Jaeger 1986; Wang & Derwing 1986);
- unproductive if they are not represented with the same spelling, even though they historically come from the Great Vowel Shift (McMahon 2007).

¹³⁸ That may be less of an issue in Substance-Free Phonology, in which ‘unnatural’ or ‘crazy’ processes are not a problem because phonological primes need not be grounded in phonetics.

Thus, if we assume that the vowel system is centred around two series of vowel (leaving r vowels aside), one ‘short’ (or monomoraic, sometimes called ‘checked’ and written \check{V}) and one ‘long’ (or bimoraic, sometimes called ‘free’ and written \bar{V}), then orthography builds connections between the vowels that alternate in morphologically related words and have the same spelling, as shown in (100). The graphophonological grammar then captures the regularities in the distribution of each type of vowel.

(100)	/a/	↔	<a>	↔	/ɛj/
	/ɛ/	↔	<e>	↔	/ij/
	/ɪ/	↔	<i,y>	↔	/aj/
	/ɔ/	↔	<o>	↔	/əw/
	/ə/ə/	↔	<u>	↔	/(j)uw/

Likewise, *SPE*’s rule of Velar Softening, illustrated by the examples in (101), has been treated as ‘unnatural’ (Myers 1999; Pierrehumbert 2006) and rejected as a phonological rule (Hyman 1975; J. Kaye 1995).

(101)	criti/k/	◆	criti/s/ize
	electri/k/	◆	electri/s/ity
	alle/g/ation	◆	alle/dʒ/e
	analo/g/	◆	analo/dʒ/y

Here, a similar analysis could be developed, in which /k/ and /s/ are connected through the spelling <c> and /g/ and /dʒ/ through /g/. As suggested by Montgomery (2001: 226) and Raffelsiefen (2010: 197), the regularity in the alternations can be captured by referring to orthography with a rule along the lines of (102).

(102)	a.	<c> → /s/	/ __ <{i, e, y}>
		<g> → /dʒ/	
	b.	<c> → /k/	elsewhere
		<g> → /g/	

Thus, it can be argued that such problematic alternations can be captured if orthographic information is taken into account.

4.3.2. Alternations requiring information provided by orthography

There are alternations for which the phonology alone cannot predict the generation of novel forms. One first instance of this are alternations between full vowels and schwa such as *atom* /'atəm/ ◆ *atomic* /ə'tɒmɪk/ (Giegerich (1992, 1999; Guierre 1979; Montgomery 2001, 2005). If a language user knows *atom* and has to generate *atomic* for the first time, they cannot know which vowel should be used in the second syllable, because *-ic* requires it to be stressed and many vowels may alternate with schwa (e.g. *brutal* /'brʌwtəl/ ◆ *brutality* /brʌw'tæləti/, *totem*

/ˈtəwtəm/ ♦ *totemic* /təwˈtɛmɪk/, *autumn* /ˈo:təm/ ♦ *autumnal* /o:ˈtəmnəl/. One could assume, as Anderson (1981) does, that they will first assume that *atom* has a melodically blank second vowel and restructure its underlying representation to |atəm| only when *atomic* is learned. However, if they know that *atom* is spelled <atom> and know that <o> is realised /ə/ when followed by <Cic#> (e.g. *alcoholic*, *chaotic*, *economic*), the pronunciation of *atomic* can be generated without it being heard, and without restructuring the underlying representation of *atom*. A small survey by Giegerich (1992) shows that speakers regularly use orthographic information if asked to generate novel forms which bear stress on vowels that are reduced in existing names (e.g. *Anders*/ə/n → *Anders*/əw/nian).

Another case in which the phonology does not have access to the information required to properly derive new words from their base is words in which a cluster containing a nasal alternates with a base-final single nasal (103).

(103) a.	damn /ˈdam/	♦	b.	damnation /damˈneɪʃən/
	bomb /ˈbɒm/	♦		bombard /bɒmˈbɑ:d/
	long /ˈlɒŋ/	♦		longer /ˈlɒŋgə/
	sign /ˈsaɪn/	♦		signature /ˈsɪɡnətʃə/
	paradigm /ˈpærədəɪm/	♦		paradigmatic /ˌpærədəɪɡˈmætɪk/

Once again, a language user who has not encountered the forms with the cluster cannot generate those forms from their morphological bases if they do not know their orthographic representation.¹³⁹ Thus, Giegerich (1999: 130) proposes an orthographically-guided rule to account for such alternations. I assume that the alternations in (103) can be captured in two ways, both assuming that the phonology has a phonotactic constraint banning base-final clusters involving a nasal and that orthographic information is available:

- Assuming that learning the spelling of a word may add a consonant to that word's underlying representation, although it is systematically banned base-finally on the surface (e.g. *damn* would be /damn/) ¹⁴⁰;
- Assuming that consonants may be inserted for the derivatives of bases that end in a nasal and are spelled with more than a consonant letter representing a nasal (e.g. <ng>).

In the cases that I have just reviewed, orthographic representations provide information that is unavailable in the phonology alone.

¹³⁹ Note that language users readily do the opposite and can generate noncewords such as [ˈlɪm] and [ˈlɪmɪŋ] from [lɪmˈneɪʃən] rather than [ˈlɪmn] and [ˈlɪmnɪŋ] (Mohanani 1986: 23).

¹⁴⁰ This kind of restructuring is suggested for French learners by Chevrot (1998), who reports that *parce que* ‘because’ is realised [pɑskə] (which is the most common form in spoken French) by pre-literate children, but that realizations as [pɑʁskə] emerge when children learn how to read and write. He suggests that they have restructured their phonological representations so as to include /ʁ/, under the influence of the orthographic representation.

4.4. Where does orthography stand?

In my 2023 chapter on the role of orthography in English phonology, I briefly mention some of the theoretical options to account for the processes discussed in the previous section and where orthography might stand in the grammar, especially with regards to the phonology.

First, let us assume a modular framework, such that orthography and phonology are two different computational systems (modules), each with its own proprietary vocabulary. This is what Fodor (1983) calls “domain specificity”. If we take seriously the claim that learning orthography restructures phonological knowledge, then the question is how it does so. As was seen previously, one possible restructuring is that the learning of orthography builds connections between vowels that share their spellings (e.g. <i>- /{ɪ, əj, ə:, əjə}/</i>), although this certainly does not override natural classes that would have been learned before orthography (e.g. vowels sharing a feature/element). Then, “anomalies” in the distribution of monomoraic and bimoraic variants may become regularized through other pieces of information in the spelling. For instance, the generalization that short vowels are to be found before consonant clusters (or in closed syllables) may be used to capture the presence of short vowels in words such as *vanilla*, *grammar* or *colossal* if the orthographic geminate is translated in some way into the phonology (e.g. it inserts a mora in the coda position that is not attached to any melodic content). As mentioned in the previous section, knowing the orthographic representation of a word may lead the learner to insert additional segments in the phonological representation of that word, especially in cases in which segments are banned by phonotactic restrictions (e.g. in cases of final clusters involving a nasal; see (103a)) or frequently undergo elision (e.g. *family* [ˈfam(ə)lɪj]). Learning orthographic forms may also serve to establish the phonological representation of words in which neutralization occurs. For example, in North American varieties of English, flapping may neutralize the contrast between /t/ and /d/. In certain cases, alternations may help the learner establish what is the source of surface [r] (e.g. *write* [ˈɹaɪt] ~ *writer* [ˈɹaɪrə]; *ride* [ˈɹaɪd] ~ *rider* [ˈɹaɪrə])¹⁴¹, but that may not be feasible in cases where there are no alternations (e.g. *spider* [ˈspajrə]). In such cases, it is possible that learners have a consonant that is underspecified for voice (or some form of archiphoneme) until its orthographic form is learned and disambiguates the phonological representation to /t/ or /d/.

A second option would be to assume that learning orthography creates a hybrid level of representation in which there are elements from both phonology and orthography. For example, for a word such as *electric*, the different levels of representation may be as in (104).

(104)	Phonology	/ɪ ˈl ɛ k t ɹ ɪ k/
		⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮
	Graphophonology	E l E C t ɹ I C
		⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮
	Orthography	<e l e c t r i c>

¹⁴¹ The learner is indeed quite likely faced with the difficulty of establishing what the underlying phoneme for [r] is, as flapping is almost systematically realized where relevant in both child-directed speech and adult-directed speech (Fritche *et al.* 2021). In some cases, the vowel may indicate what the underlying consonant is (e.g. because of raising or clipping) and may thus be used to establish the phonological representation of that consonant, as suggested by Bermúdez-Otero (2003).

In the graphophonological representation in (104), capital letters are shorthand for series such as /{ɪ, ʌj, ə:, ʌjə}/, whose specific realization is determined by context. Such representations could then be used to compute phonological representations (if they are unknown) in a way that is very similar to what is done using the abstract underlying representations posited in *SPE* or Jensen (2022), with the notable difference that units such as E are not assumed to have any features. However, I am unsure that such a level is required, as it mostly encodes information from the spelling, simply highlighting graphemes that have variable realizations.

Thus, another option is to assume that orthographic elements may intervene in phonological computation. This is the option taken by Giegerich (1999) to capture cases like those in (103). He proposes the orthographically-guided rule in (105) to account for cases in which an /n/ is inserted after a base-final /m/ in words ending in <mn>.

(105) *n-insertion* (taken from Giegerich 1999: 130)

$$\emptyset \rightarrow [n] / [m] \text{ ______ } \begin{array}{c} \text{<mn>} \\ | \\ \text{V} \end{array}$$

A more recent and more extensive model implementing this idea is the extension of the BiPhon model (for “Bidirectional Phonetics and Phonology”; Boersma 2007, 2011) proposed by Hamann & Colombo (2017). This model is an OT model in which different families of constraint apply in parallel on one or between different levels of representation. Six levels of representations are used: two semantic levels, two phonological levels and two phonetic levels. The phonological and phonetic levels and the family of constraints connecting them are shown in Figure 39.

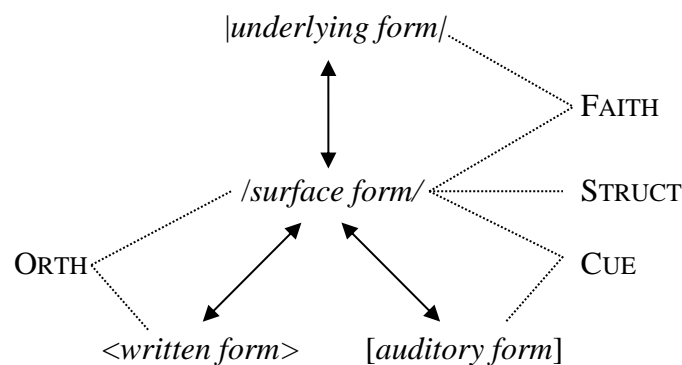


Figure 39. The BiPhon model extend so as to include orthographic representations (Hamann & Colombo 2017)

The model is said to be bidirectional because one of its central assumptions is that the same constraint hierarchy applies in production and in perception, unlike most phonological models, which focus on production alone. The revised model in Figure 39 incorporates ORTH constraints which regulate the mapping between orthographic forms and surface phonological representations. For example, one type of constraint that they propose is <ɣ>/P/, which states

that a violation mark should be assigned “to every grapheme $\langle\gamma\rangle$ that is not mapped onto the phonological form /P/ and vice versa”. Such a constraint that they use in their analysis of Italian is $\langle t \rangle / t /$, which requires that $\langle t \rangle$ be realized /t/ and vice versa. Note that, as pointed by Hamann (2020), constraints that are necessary in the phonology anyway need not be replicated in the reading grammar (e.g. the ban on base-final /mn/ clusters or the absence of geminates in words with nonanalytic morphology). Hamann & Colombo also follow the psycholinguistic literature by assuming a dual-route model of reading in which ORTH constraints represent the “sub-lexical” route. The other route is called the “lexical” route, and it consists in mapping a whole-word’s orthographic representation onto a whole-word phonological (and semantic) representation. This whole-word access is particularly necessary in the case of exceptions to the generalizations of the sub-lexical route or in the case of homographs (e.g. *tear* /'tɪ:/ or /'tɛ:/). However, the model has been only applied to languages with relatively shallow orthographic systems, such as Italian or German, and it remains to be seen how it can be extended to a more complex system such as English. In English, we might assume that the learner starts with simple mappings such as $\langle t \rangle \leftrightarrow /t/$ and then proceeds to more complex cases, such as the series represented by the different graphemes (e.g. $\langle a \rangle - / \{a, \epsilon j, \alpha:, \epsilon:\} /$, $\langle c \rangle - / \{s, k\} /$). They may then acquire local generalizations such as “ $\langle a \rangle$ is realised /a/ before two consonants” and eventually more global generalizations such as “vowel monographs are read as their short/monomoraic realization before certain orthographic consonant clusters”. As shown by the studies on reading (e.g. Treiman *et al.* 2020; Treiman & Kessler 2023), such generalizations are not categorical but probabilistic. Let us take very simple examples of how the adult system of a literate speaker might be formalized when it comes to reading. Let us assume the constraints in (106), assumed to be unranked, so that we can derive the different realizations of $\langle o \rangle$ in the words in (107). I am leaving out the realization /o:/, which is only found in the presence of $\langle r \rangle$.

- (106) a. $\langle o \rangle / \{ \text{ɔ}, \text{əw}, \} /$
 $\langle o \rangle$ should be realized as /ɔ/ or /əw/.
- b. *NN
 No adjacent nuclei.
- c. FTBIN
 Feet must be binary at some level of analysis (μ , σ).

(107)

		<o>/{ɔ, əw }/	*NN	FTBIN
a.	<so>	☞ (səw)		
		(sɔ)		*
		(sɔj)	*	
b.	<pot>	(pəwt)		*
		☞ (pɔt)		
		(pat)	*	
c.	<poem>	(pɔ.əm)	*	
		(pɔ)əm	*	*
		☞ (pəw)əm		
		(paw)əm	*	

In (107a), (səw) is selected because it is the only candidate that does not violate any constraints: (sɔ) violates FTBIN and (sɔj) violates the ORTH constraint <o>/{ɔ, əw }/. Likewise, in (107b), (pɔt) is selected because it does not violate any constraint, unlike the other two candidates. In (107c), it is assumed that the presence of a diphthong is favoured here because it avoids the hiatus found in other candidates (as was suggested in §2.1.1).¹⁴²

The details of the analysis need to be worked out as the system is very complex. I have tried to devise such a global analysis which would capture stress and vowels, but there are still many points that need to be solved. One of the issues that the theory would also have to solve is how the lexical route is implemented. One possible solution would be to employ mechanisms that are comparable to morphological blocking at the stem-level in Stratal Phonology (see §3.2.2), so that lexical exceptions are preserved. This could be achieved through an ORTH constraint requiring faithfulness to the whole-word phonological representation stored in the lexicon (let us simply call it FAITHWD here), which would be ranked higher than other constraints.

(108)

		FAITHWD	<o>/{ɔ, əw }/	*NN	FTBIN
<post>	(pəwst)				*
	(pɔst)	*!			
	(past)	*!	*		

For example, in (108), FAITH preserves the lexically stored representation for the word *post* and prevents the regular application of the other constraints, which would otherwise generate /'pɔst/. This has the effect that, once a language user knows both the orthographic and the phonological representation for a word, the remaining ORTH and phonological constraints function essentially in lexical redundancy mode, i.e. they mostly capture the statistically dominant patterns found in the lexicon. They may be used generatively only if the language user is presented with the orthographic representation of a word that they do not know.¹⁴³ A further complication is divisions in the grammar that may be necessary even if one ignores orthography (strata,

¹⁴² The candidate (pə.wə)m is not evaluated here because it is assumed that glides syllabify in the coda.

¹⁴³ However, violating ORTH or markedness constraints to satisfy FAITHWD may have some processing 'cost', as shown by the fact that lexical exceptions tend to be read slower than regular words (see e.g. Rastle & Coltheart 2000).

sublexica) should also be incorporated in the model. For example, vowel alternations such as those seen in (99) are specific to what Stratal Phonology calls the stem-level and not those of the word-level, and spelling-to-sound correspondences may differ depending on the type of vocabulary involved (see Ch. 5).

Finally, let us conclude this section with one possible effect of the inclusion of orthographic information into phonological computation which I had not considered before. Alongside ORTH constraints, we might have a constraint forbidding the realization of sounds that do not have orthographic counterparts. In other words, from the perspective of a reading grammar, a constraint against insertions. Following common practice in OT, let us call it DEP. Such a constraint could explain why (at least according to Wells (2008)), it would be more common to have a stop between the nasal and the voiceless fricative in the words in (109a) than in the words in (109b).

(109) a.	lynx	[ˈlɪŋ(k)s]	b.	Kingston	[ˈkʰɪŋ(k)stən]
	Sampson	[ˈsəm(p)sən]		Samson	[ˈsəm(p)sən]
	antsy	[ˈan(t)sɪj]		sense	[ˈsen(t)s]

Likewise, DEP could be partly responsible for the difference in the rates of linking r (e.g. *doo*[ɹ] *outside*) and intrusive r (e.g. *law*[ɹ] *enforcement*). The former, in which [ɹ] can be mapped to an orthographic <r>, is more common than the latter, in which there is no orthographic <r> (Durand *et al.* 2015; Mompeán-Gonzalez & Mompeán-Guillamón 2009). Although this is clearly not the only factor at play, orthography may interfere with phonological factors such as hiatus avoidance and resyllabification. It is quite likely that the weight attributed to DEP is sociolinguistically conditioned, for example depending on language users' relationship with prescriptivism.

4.5. From secondary stressed vowels to a possible category-specific vocalic behaviour

4.5.1. The value of secondary stressed vowels

In my dissertation (Dabouis 2016b), one of the aspects that I studied regarding pretonic syllables that carry secondary stress concerns which vowels are found in syllables in which the vowel is followed by a single consonant. Previous research in the Guierrian School conducted by J.-M. Fournier (1994) and Guierre 1979) had shown that vowels with secondary stress overall follow the same generalizations as vowels with primary stress. Those two studies indeed show that digraphs are realized as \bar{V} s (e.g. *òubliète*, *màusoléum*, *èucalýptus*), that having zero or over two consonants after the stressed vowel is associated to the same vocalic behaviour, i.e. \bar{V} s before zero consonants (e.g. *crèativité*, *Làodicéan*) and \check{V} s before a consonant cluster (e.g. *kàngaróo*, *dìplomátic*, *àllegréto*, *èxploítation*), unless the first consonant is <r>, in which case we have \check{V} s (e.g. *jàrgonéll^re*, *gòrgonzóla*, *pèrpetúity*). Finally, if the vowel is <u> followed by a single consonant, it is systematically a \bar{V} (e.g. *fùsilláde*, *ùnánimíty*, *mùtinéer*). Some environments that are associated to a stable vocalic behaviour are simply impossible for

pretonic secondary stressed vowels, such as being followed by certain final sequences in the following syllable (e.g. -V#, -ic(s)#).

However, there is disagreement between the two authors regarding the behaviour of the pretonic sequence (i.e. the syllables that precede primary stress), which has an impact on the interpretation of rules that refer to the position of the stressed vowel. Three such rules are concerned:

- **The “CiV” rule:** for primary stressed vowels, that rule yields \bar{V} s if the vowel is not <i, y> and is followed by the sequence -{i, e}+V(C₀(e)). J.-M. Fournier assumes that that sequence should be word-final, but Guierre does not and so that would predict that the rule could apply if that sequence follows the secondary stressed vowel, even if it is not final.
- **Trisyllabic Shortening** (or Laxing, also called “Luick’s rule” in the Guierrian School): that rule states that the stressed vowel should be a \check{V} if the stressed vowel is followed by two vowels. Guierre and J.-M. Fournier’s approaches predict different behaviours as Guierre counts only syllables in the pretonic sequence and J.-M. Fournier counts all syllables that follow the stressed vowel. Crucially, Guierre does not mention what should happen if there are two syllables in the pretonic sequence (e.g. *vòluntéer, màcaróon, rècomménd*) while J.-M. Fournier predicts that Trisyllabic Shortening applies in that environment.
- **The Rule of the Prefinal:** that rule states that stressed vowels should be \bar{V} s if they are in the penult. Like Trisyllabic Shortening, Guierre and J.-M. Fournier’s approaches make different predictions: J.-M. Fournier’s approach makes this rule inapplicable while Guierre’s predicts that it applies in words such as *chiàroscúro*.

Moreover, J.-M. Fournier (1994) shows that, in suffixal derivatives in which primary stress is further rightwards than it is in the base, secondary stressed vowels have the same value as that found in the base, regardless of the position of secondary stress. This can be shown by the presence of \bar{V} s in positions in which Trisyllabic Shortening would be expected to apply. This goes against the assumptions present in the Guierrian School at the time according to which the value of the vowel was to be ‘preserved’ in cases in which the position of secondary stress can be argued to depend on the position of primary stress in the base (110a), but not if the position of stress is constrained by rhythmic principles (110b), i.e. if there are two pretonic syllables, secondary stress is normally found on the first syllable, in both non-derived and derived words.

(110)	a.	sòcialístic	◆	sócial	b.	màjorétte	◆	májor
		mèdiùmístic	◆	médiùm		privatéer	◆	prívate
		supèriórité	◆	supériór		nòtabílité	◆	nótable
		desirábilité	◆	desírablé		fátalístic	◆	fátal

Thus, in my dissertation, I tested those hypotheses and, more generally, sought to establish what the regular value for secondary stressed vowels that are followed by a single consonant and which are neither a digraph nor <u> is. In non-derived words (monomorphemic words or words derived from bound roots), the results were clear: \check{V} s are found in about 90%

of words whenever there is at least one syllable after the secondary stressed vowel. Having more than one syllable afterwards or having a ‘CiV’ sequence appears to have no effect.¹⁴⁴ This is consistent with Fournier’s approach rather than Guierre’s. However, a new result is the behaviour of secondary stressed vowels in the initial pretonic position. Although few words are concerned, those vowels are realised as \bar{V} s (including foreign free vowels \bar{V}^f ; see §5.1) in over 90% of cases (e.g. *bètise*, *dèmentia*, *hòtél*, *migráte*). This is what I have called the ‘Rule of the Initial Pretonic’.

In stress-shifted suffixal derivatives in which the relevant vowel in the base is reduced (e.g. /ə/*cademy* ♦ /à/*cadémic*), the results are less clear-cut as there is a slightly lower proportion of \check{V} s (around 75%), probably attributable to a quantity interaction with the second syllable akin to the Arab Rule (see §2.2.1). If the base contains a full vowel in the relevant position, the secondary stressed vowel in the derivative has the same value in 99% of cases, be it \bar{V} (111a), \check{V} (111b) or variation between those two values (111c).

- | | | | | |
|-------|----|--------------------|---|------------------|
| (111) | a. | neg/əw/tiátion | ♦ | neg/əw/tiate |
| | | /éj/miábility | ♦ | /éj/miable |
| | b. | c/à/tegorizátion | ♦ | c/á/tegorize |
| | | fam/i/liárity | ♦ | fam/i/liar |
| | c. | am/ij ~ è/nábility | ♦ | am/ij ~ é/nable |
| | | exc/ò: ~ ò/riátion | ♦ | exc/ó: ~ ó/riate |

It is also insightful to note that all the derivatives in which the base has a \check{V} also have a \check{V} (as in (21b)), and that whenever there is a mismatch between the two, it nearly always involves the use of a \check{V} in the derivative that is not found in the base. Those two elements further confirm that the default value of the vowel in those positions is \check{V} .

4.5.2. A foot-based perspective

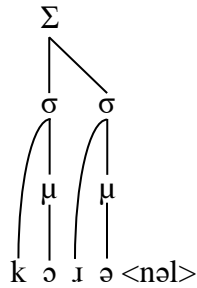
In 2018, I published an expanded version of that part of my dissertation in *Anglophonia*. One of the significant additions was the use of foot structure to analyse the results reported in the previous section. As was seen in §2.1.2.2, analyses of English phonology using Prosodic Phonology often assume that the default foot type is the bimoraic trochee. This means that a foot should either be formed of two monomoraic syllables or one bimoraic syllable.

Generative analyses usually assume that vowels are underlyingly specified and have, since Hayes (1982), embraced the idea that certain word-final constituents are ‘extrametrical’, i.e. invisible to the mechanism building prosodic structure. In nouns and for certain suffixes (e.g. adjectival *-al*), the final syllable is assumed to be extrametrical, while only the final consonant of verbs is extrametrical. Following those assumptions, the contrast between (112a)

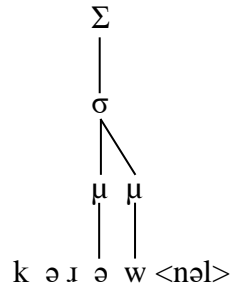
¹⁴⁴ The exceptions are mainly foreign words with foreign spelling-to-sound correspondences. If those are excluded, the proportion of checked vowels goes up to 99%.

and (112b) may be assumed to stem from different underlying representations, with an underlying short vowel in the former but an underlying long vowel in the latter.

(112) a. córonal



b. corónal



Another way of looking at the vocalic variation in (112) is to assume, following what is usually done in the Guierrian School, that the value of the vowel depends on the position of stress. Thus, the penultimate vowel is reduced in (112a) because it is not the head of the foot and is long in (112b) because it is the foot head and because feet should be bimoraic trochees (the final syllable being extrametrical). This means that what I have referred to as ‘Trisyllabic Shortening’ and the ‘Rule of the Prefinal’ can both be captured using the same restriction of foot structure, provided that there is an independent constraint determining where foot heads should be. Assuming that the final syllable is extrametrical means that Trisyllabic Shortening does not require two syllables after the stressed syllable but only one and should be reanalysed as ‘Trochaic Shortening’, as proposed by Prince (1991).

Returning to the value of vowels with secondary stress, if what conditions vowel values is the moraic trochee, then the presence of \check{V} s when the following vowel is unstressed is entirely expected. Indeed, such vowels are in the same environment as vowels that are in the environment for ‘Trisyllabic’ Shortening, as in (112a):

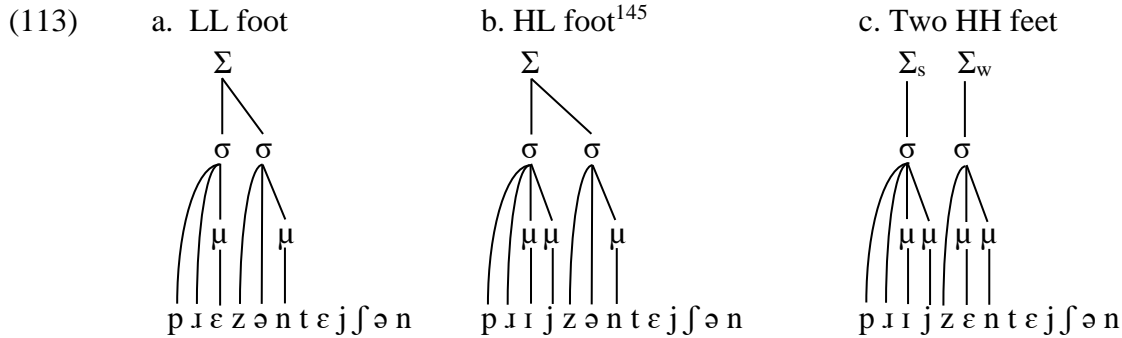
$$(\acute{\sigma} \sigma) \langle \sigma \rangle = (\acute{\sigma} \sigma) (\acute{\sigma})$$

The analysis also predicts the Rule of the Initial Pretonic as the initial syllable, if it is to form a foot on its own (because the following syllable belongs to the head foot), should have a \bar{V} to reach bimoraicity, just as in (112b).

This approach entails that a word such as *mètricálicity*, which has a short vowel in its first syllable, cannot be analysed as a case of cyclic overapplication of ‘Trisyllabic Shortening’ (as suggested by Bermúdez-Otero (2013) if it is analysed as Trochaic Shortening. Under that analysis, the first vowel in *mètricálicity* is in the right environment for shortening (just as the first vowel of *sòlidálicity* is) and does not show cyclic overapplication. However, the first vowel of a word like *nòtability* does show an underapplication of Trochaic Shortening that may be attributed to the fact that that vowel was not in the right environment for Trochaic Shortening in its base, *nótable*. Thus, the vowel of *mètricálicity* cannot be used to demonstrate ‘cyclic’ effects, but that of *nòtability* may. As for a word such as *fàtálicity* (also mentioned by Bermúdez-Otero (2013)), it cannot be compared to either of those words because its secondary stressed

vowel is in the initial pretonic environment, in which long vowels are the norm when vowels are stressed.

In the case of derivatives in which the vowel corresponding to the one that bears secondary stress is unstressed in the base, it can be assumed that the variation observed is due to various footing options for the first two syllables which interact with faithfulness to the base, shown in with the example of *presentation*.



Option (113a) would adapt the derivative to the dominant pattern found in non-derived words at the cost of reduced faithfulness to the base *present* /pɪə'zɛnt/, while option (113c) can be seen as resulting from that faithfulness: keeping the foot on the second syllable while at the same time seeking to avoid a strong clash with the head foot requires another (bimoraic) foot on the first syllable. Option (113b) is more difficult to account for, maybe except if it is seen as an intermediary stage between (113c) and (113a).

That analysis makes further predictions beyond pretonic vowels with secondary stress, and it leads to interesting observations regarding the phonological behaviour of different syntactic categories.

4.5.3. Vowel values and syntactic categories

Bermúdez-Otero & McMahon (2006) develop an analysis in Stratal OT which assumes that feet are binary in English, and they note that Trochaic Shortening applies differently to different classes of words. They capture those differences by assuming that those different classes obey different cophonologies, in which the constraints enforcing the parsing of syllables into feet and the one enforcing the extrametricality of the final syllable have different rankings. In their analysis, all of the cases in (114) abide by Trochaic Shortening, but those in (114a) have an extrametrical final syllable while those in (114b) only have an extrametrical final consonant.

(114) a.	Nouns:	A(měri)<ca>	
		sin(cěri)<ty>	cf. sincēre
	Regularly suffixed adjectives:	(nātu)<ral>	cf. nāture

¹⁴⁵ I am here assuming that unstressed /ə/ is moraless. An alternative would be to assume that /n/ occupies the nucleus position, but the second syllable would still be monomoraic. This option might alternatively be assumed to involve a minimal foot projection for the first syllable, and a higher foot projection to which the second syllable attaches.

b.	Verbs:	de(vělo)<p>	
	Non-derived adjectives:	de(crěpi)<t>	
	-ic adjectives:	cy(clōni)<c>	cf. cyclōne

If one takes over that analysis into the usual Guierrian perspective according to which vowel values are determined after stress, the prediction is that there should be a difference in vowel values between different syntactic categories. This is something that I have explored in Dabouis (2018). In that paper, I tested the prediction made by the assumption that verbs and nouns are subjected to different rules of extrametricality, according to which we would expect vowels in open penultimate syllables in verbs to be short, so as to form a disyllabic moraic trochee. I used the data from Dabouis & J.-M. Fournier (2023) and found that, if we leave out all the verbs that are subjected to rules that are hierarchically superior to Luick's rule in J.-M. Fournier's (2010b) system, we find short vowels in 62/69 (89%) of cases (e.g. *bóther*, *fáthom*, *imáagine*). This could suggest that there is indeed a specific generalization that applies to verbs with penultimate stress, and this is an observation that has also been made previously by researchers working on the distribution of short and long vowels, such as Ross (1972) or Dickerson (1980).

However, it is possible that this predominance of short vowels among verbs has nothing to do with syntactic category as such. Nessly (1982) challenges Dickerson's (1980) analysis, which assumes that penult-stressed verbs have short vowels because they are verbs. In the sample provided by Dickerson, he notes that there is an overrepresentation of words in which <v> follows the stressed vowel. As this consonant letter is only rarely doubled, Nessly assumes that this could entail a greater likelihood for vowels followed by <v> to be short. Supporting evidence for this claim comes from a recent study from Treiman & Kessler (2023), which shows that, among disyllabic words with initial stress and a single medial orthographic consonant, <v> is the consonant associated with the highest rate of short vowels (49%), and the only medial consonant that cannot double. Nessly also notes that many verbs in Dickerson's sample are prefixed (e.g. *consíder*, *delíver*, *enámor*, *inhibít*) and that we also find short vowels in prefixed adjectives (e.g. *comp/ósite*, *expl/í/cit*, *exqu/í/site*, *impl/í/cit*, *ins/í/pid*, *intr/é/pid*). Finally, he notes that the verbs in Dickerson's sample also contain endings that are particularly associated to the presence of short vowels, even in words that are not verbs (e.g. <-it> *crédít*, *dígít*, *édít*, *hábit*, *límit*, *mérít*, *prófit*, *tácit*). Treiman & Kessler (2023) also test the role of endings and indeed find that <-it> is the fifth most strongly associated with preceding short vowels in their data. While in my 2018 paper, I had concluded that there surely must be a rule that applies to verbs with penult stress, I now believe that more work is necessary in order to make any such claims, so as to make sure that the potential influence from other factors is taken into account.¹⁴⁶ Moreover, as shown by Treiman & Kessler (2023), the association between those factors and different vowel values are probabilistic, and so it is possible that, should there be a statistically significant association between syntactic category and vowel values, that association may not be anywhere near categorical. If further investigations confirm a role for endings and medial consonants, then the foot-based account described in this section will have to be revised.

¹⁴⁶ However, a rule positing that the primary stressed vowel of verbs with penultimate stress and in which that vowel is followed by a single consonant and is neither a digraph nor <u> should be a \check{V} can be useful pedagogically as it can account for very common verbs which are otherwise unaccounted for in J.-M. Fournier's model (2010b).

CHAPTER 5

Sublexica in English

5.1. Is there a single phonological system in English?

In §3.2.2, I have discussed the core principles of Stratal Phonology. In that approach, different morphosyntactic constituents are assigned to different phonological grammars that apply serially. However, there are languages for which it has been argued that, for the same type of morphosyntactic units, different subparts of the lexicon (e.g. native words vs. borrowings) follow different phonological generalizations. The idea is quite old (see Lees (1961) on Turkish and Postal (1968) on Mohawk) and one prominent case is Japanese (Itô & Mester 1995, 1999), in which it has been argued that the different sublexica are organized in a core-periphery structure, so as to capture different degrees of nativization. For English, following *SPE*, researchers have traditionally distinguished ‘Linate’ from ‘Native’ vocabulary. The authors of *SPE* also distinguished words of Greek origin, and argued that

This classification is functional in the language and must be presumed to be represented in the internalized grammar. It is justified not by the historical development of the language but by the applicability of phonological and morphological rules. (*SPE*: 373)

Those distinctions are argued to be necessary for the application of processes such as Velar Softening (see §4.3.1). Pater (1994) also argues that “light penult stress” (e.g. *anténna*, *Cincinnati*, *Gehenna*, *Vanessa*) is a property of the “periphery” of the grammar, while antepenult stress a property of the “core”, following the same sort of model advocated by Itô & Mester (1995). Distinctions between different parts of the vocabulary are often used in morphology as well. For example, Aronoff (1976: 51 ff.) notes that certain affixes only attach to Linate bases. He mainly uses the distinction between “native” and “Linate”, but notes that “[t]here are probably even further subdivisions, into *greek*, *romance*, etc.”. Giegerich (1999: §2.2.1) also argues that the “Affix Ordering Generalization” (Siegel 1974) is probably attributable to the selectional restrictions of affixes in terms of “native” and “Linate” features.

However, we also find claims such as the following: “in some languages, borrowed words exhibit rather different phonological patterns from those of native words, but this does not appear to be the case in the stress system of English” (Hammond 1999: 284), which goes against those approaches (at least regarding stress). In many reference works on English phonology (e.g. *SPE*; Burzio 1994; Hayes 1982; Pater 2000), we find loanwords being used to illustrate the stress rules of English (e.g. *Ticonderoga*, *Monongahela*), arguably because the donor languages do not have stress. Thus, as those words have structures (e.g. long pretonic sequences) that are hardly attested in English monomorphemic words, they give one the chance

to see patterns that are difficult to observe in the rest of the lexicon. I have never found any mention of the possibility that a ‘foreign’ or ‘peripheral’ grammar could apply to those words.

Within the Guierrian School, it has been common practice to distinguish between native and borrowed words. This can partly be explained by the fact that this school has been interested in graphophonological generalizations, for which distinctions between different parts of the lexicon seem absolutely necessary (Carney 1994). For example, J.-M. Fournier (2010: 113) calls the correspondences found for certain vowels in borrowed words “foreign free vowels” (e.g. <a> - /ɑː/ *armada, banana*; <e> - /ɛj/ *cliché, peso*; <i> - /ij/ *bikini, mosquito*). J.-M. Fournier (2007) goes further in claiming that English phonology, and especially its stress system, is a hybrid system, and that syllable weight only impacts stress clearly in words borrowed from Romance languages or Modern Latin. He argues that English “has inherited not *one* phonology but *several* phonologies, whose workings are determined by morphology and word-length, but also by learned or foreign word characteristics” (emphasis in the original). This is assumed to be a result of the history of English, which results from a blending of Romance and Germanic vocabulary (see Minkova (2006) for a similar view).

Thus, it seemed to me that it was necessary to control for possible differences between subparts of the lexicon when studying English phonology. When working on pretonic secondary stress placement for my dissertation (Dabouis 2016b), I initially classified my data into three categories: foreign, learned and the remaining words. It turned out that no statistically significant differences exist with regard to pretonic secondary stress placement between these three subsets, and so they were eventually collapsed into a single dataset. We did the same in our study of vowel reduction with Jean-Michel Fournier (see §2.2.2), which revealed an effect of foreignness: foreign words (identified on formal criteria such as endings, foreign spelling-to-sound correspondences and semantics) have reduced vowels less often than non-foreign words, independently of other factors such as syllable structure or frequency. Eventually, it became necessary to propose a more detailed model of those subparts of the lexicon, which I will call **sublexica**, and what their properties are.

5.2. Sketching a model

We have seen that the idea of polysystemicity appears to be quite widespread, but I was not aware of any formal treatments of that idea in English. In the Guierrian School, it is expressed clearly only by J.-M. Fournier (2007) with the claim that **English has several phonologies**. Therefore, I started by making this into an explicit claim of that approach in my dissertation and then in our joint paper on the Guierrian School with Jean-Michel Fournier, Pierre Fournier and Marjolaine Martin (Dabouis *et al.* 2023). It also seemed necessary to go further and formalize a model specifying which sublexica exist in English and what their properties are.

This is what we set out to do with Pierre Fournier in a chapter published in a collective volume entitled *Models and Modelisation in Linguistics* (Dabouis & P. Fournier 2022). In that chapter, we adopt the terminology developed by Carney (1994) for English spelling-to-sound correspondences, and identify four main sublexica, which we define as follows:

- §Core: most of the vocabulary, with regular spelling-to-sound correspondences, nonspecialised semantics and productive morphological processes (e.g. affixation by juxtaposition, compounding). It will sometimes be necessary to distinguish two subsystems within §Core, although the two are very much intertwined:
 - §Core-Native: words of Germanic origin, which constitute most of the basic vocabulary and are usually quite short.
 - §Core-Latinate: old borrowings from Latin or French and which may not be clearly identified as such. Many of them are quite frequent and words tend to be longer than those of §Core-Native.
- §French: words associated to French pronunciation (e.g. final stress) and spelling-to-sound correspondences. They often carry semantics associated to a form of “prestige”.
- §Foreign: words borrowed from or associated to Romance languages such as Italian, Spanish or Portuguese (to the exclusion of French) but it may also include words from other language families such as Russian or Japanese. These are massively stressed on their penultimate syllable and often refer to foreign cultures (e.g. food, clothes, traditions).
- §Learned: technical or scientific vocabulary, often borrowed directly or constructed from elements borrowed from Greek or Latin. These include neoclassical compounds.

Dabouis & P. Fournier (2022: 217-218)

In that paper, we do not provide new data, but we discuss many existing observations on stress, graphophonology, segmental structure, morphology and semantics that can be organized into a coherent system. We show that those sublexica present a cluster of properties on different levels. For example, items belonging to §FRENCH often have final stress, they have some specific spelling-to-sound correspondences (e.g. <i> - /ij/; <ou> - /ʊw/; <gn> - /nj/) and they tend to have semantics associated with an idea of “prestige” (Chadelat 2000). I reproduce the summary of the different properties of each sublexicon that we identified in Table 25.

Table 25. Summary of the properties of the different sublexica in English (adapted from Dabouis & P. Fournier (2022: 236))

		Stress		Graphophonology		Morphology	Semantics	Segmental structure
§CORE	§NATIVE	Normal Stress Rule	Neutral suffixes?	Basic correspondences	No Velar Softening	Affixation by juxtaposition	Basic vocabulary	Short words
	§LATINATE	Germanic Law	Stress-affecting or mixed suffixes?		Velar Softening	Compounding Conversion		
§FRENCH		Final stress	Regular secondary stress rules	<a> - /ɑ:/ <e> - /ɛj/ <i> - /ij/	<ch> - /ʃ/ <qu> - /k/ <ou> - /ʊw - ɔ:/ <(e)au> - /əw/	Specific suffixes (e.g. <i>-ette, -aire</i>)	« Prestige »	Distribution of /ʒ/
§FOREIGN		Some stress-affecting suffixes (e.g. <i>-ette, -aire, -eur</i>)			<c(c)> - /tʃ/ <sch> - /ʃ/ <z> - /ts/			
§LEARNED		/-10/ if prefinal consonant cluster	Specificities for secondary stress placement in suffixed neoclassical compounds?	<a> - /ɑ:/ (marginal)	<ae, oe> - /ij/ <eu> - /juw/ <ph> - /f/ <ch> - /k/	Specific suffixes (e.g. <i>-itis, -osis -us, -on</i>) Base truncation more common Neoclassical compounding	Technical, scientific or learned vocabulary	Polysyllabic words CVC Structure for B roots in neoclassical compounds

Let us mention two key features of the sublexica as we understand them. First, they are by no means waterproof. Indeed, we expect considerable variation in the extent to which they are developed for individual language users, and so a particular item may be categorized differently from one user to the next, possibly resulting in different phonological properties and thus inter-speaker variation. This view is consistent with Minkova (1997), who also adopts a polysystemic view of the English phonological system:

Following the model proposed by Itô and Mester (1995), the [Middle English] lexicon can be stratified into a core consisting of native and already assimilated words for which all relevant constraints obtain, and increasingly peripheral lexical strata mapped like concentric circles around the core, for which the core constraints are progressively weakened. The boundaries between all strata allow fluctuation”.

Second, although at least some of the sublexica are influenced by other languages, they stand in an indirect relationship with those languages. For example, in our model, most words that are borrowed from Italian would be perceived as belonging to §FOREIGN. In that sublexicon, the default position of primary stress is the penultimate syllable, a position that is also common in Italian but not systematic. Thus, our model predicts that English speakers will more often shift stress to the penult than they will to other positions. This is precisely what happens (P. Fournier 2018, 2021), and there is some evidence that speakers with different levels of proficiency in Italian assign stress in different ways: those with a lower proficiency overgeneralize penultimate stress even more than more proficient speakers. In the case of §LEARNED, the source languages (Latin and Greek) being dead, the relationship is even more indirect. Thus, we proposed a representation of the whole system, that I reproduce in Figure 40.

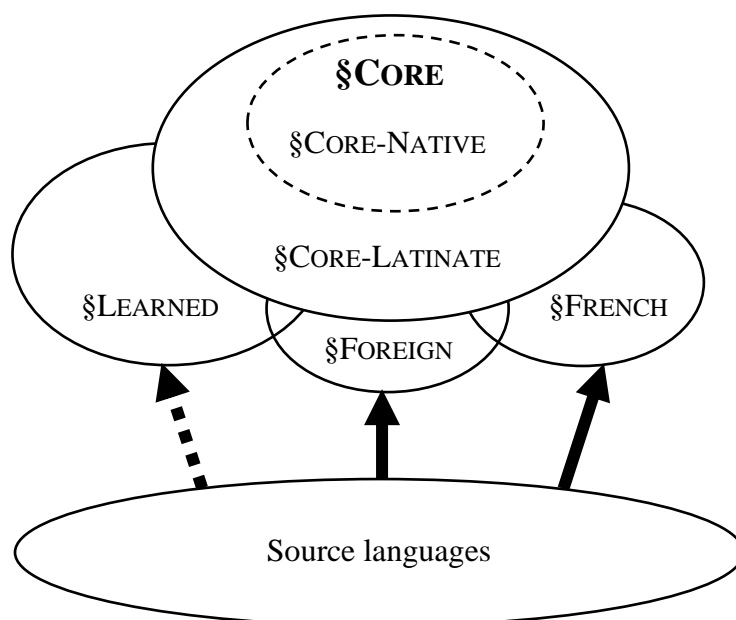


Figure 40. The sublexica of English (taken from Dabouis & P. Fournier (2022: 248))

Let us briefly comment on that representation. First, we can note that the different sublexica are represented as partially overlapping. Indeed, as was seen in Table 6, some

properties are shared by several sublexica (e.g. the correspondences <a> - /ɑ:/, <i> - /ɪj/ and <e> - /ej/ are found in both §FRENCH and §FOREIGN) and we assume that, if no characteristic of one of the more peripheral sublexica are detected, words will be assumed to belong to §CORE as a default. In the case of §CORE-NATIVE and §CORE-LATINATE, which are the two sublexica most commonly used in the literature, we assume that the overlap is very extensive and that those systems differ only marginally. Second, we represent the relationship between sublexica and the corresponding source languages with arrows, and the dashed line connected to §LEARNED is meant to represent its weaker connection to the source languages.

Finally, we formulated four predictions made by that model and which will have to be tested in future research. Those are shown in (115).

(115) Predictions made by the Dabouis & P. Fournier (2022) model

- a. If speakers are presented with words with identifiable characteristics from one of the sublexica, they will tend to use the generalizations from that sublexicon. In relevant cases, these may differ from those of §CORE.
- b. Speakers with a higher proficiency in a language related to one of the sublexica (e.g. Italian and §FOREIGN) are more likely to diverge from the default patterns of that sublexicon, and speakers with a lower proficiency will be more likely to conform to the generalizations of the sublexicon.
- c. As we can assume that a sublexicon may emerge only if speakers have been exposed to a sufficiently large number of words of that sublexicon, we can expect speakers with a large vocabulary size to conform to the generalizations of these sublexica, even when they contradict those of §CORE, whereas speakers with a smaller vocabulary will tend to conform to §CORE.
- d. Mismatches between the pronunciation of a borrowing in English and its pronunciation in its source language are more likely to arise when there is a conflict between the patterns of the borrowing and the generalizations of the associated sublexicon (and not simply those of §CORE).

Furthermore, the psychological reality of sublexica could be investigated in psycholinguistic research. There is one study which has shown that language users are aware of at least some distinctions in the spelling system. Treiman *et al.* (2019) showed participants pairs of nonwords which contained spellings that were consistent with what they call the “Latinate” system and others that mixed that system with “basic” spellings and asked them to select the more wordlike item from the pair. Their results showed that participants preferred nonwords with consistent spellings in over 98% of the time, which strongly suggests that they were aware of the existence of those orthographic subsystems. Further experiments could be conducted to test the associations between different characteristics across different sublexica. For example, we could design a lexical decision task in which we would present participants with pseudowords with consistent or inconsistent characteristics (e.g. orthography, stress) based on what was identified in Table 25. We could also conduct production experiments in which different groups of participants are presented with the same pseudowords but with different instructions. Some of

them may be told that the words are French, others that they are Italian, and yet others told nothing. We could then ask them to read out the pseudowords and how the instructions that they received affected their productions. For example, our model would predict that they would produce more penultimate stress if the pseudowords are presented as Italian or Spanish, more final stress (primary or secondary) if they are presented as French, and that we will get more foreign free vowels in pseudowords presented as being foreign than in pseudowords that are not. Another option could be to provide participants with definitions of the pseudowords that may or may not refer to foreign culture or technical or scientific concepts.

At this stage, we have not conducted studies that are specifically designed to test out those predictions or studies such as the ones I have just sketched, but rather conducted studies which explore the characteristics of the sublexica, mainly using dictionary data. Those are discussed in the next section.

5.3. Exploring the model

5.3.1. *The French sublexicon*

In 2020, I was asked to review a paper on French loanwords for the *Journal of Linguistics*, and I quickly realized that the author was Pierre Fournier. I sent a number of suggestions on the submitted version of the paper, and he quickly suggested that we work on a revision of the paper together, which I accepted. We deeply restructured his original paper and added several new analyses using statistics that were not present in the initially submitted version. After additional reviews were made, the paper was accepted for publication. It is currently a FirstView article and will appear in 2025 (Dabouis & P. Fournier 2024). In line with a lot of work that Pierre Fournier has been conducting in the past few years (see e.g. P. Fournier 2016, 2021a, 2024), the aim was to explore how loanwords are stressed in English and what factors affect the stress patterns of those loanwords. Our joint paper focuses on French loanwords, and thus may contribute to a better understanding of the §French lexicon, as those two categories largely overlap.

The study is based on a dataset of 1,043 entries¹⁴⁷ from Wells (2008) which was collected by Pierre Fournier, in which the English pronunciations are accompanied by French phonemic transcriptions. As it is often observed that French loanwords show sizeable proportions of words with final primary stress, we wanted to measure how systematic that pattern is, with a dataset that is less biased than previous studies (e.g. P. Fournier (2016) focuses only on words that contain certain French endings). Moreover, the words which we are studying are quite likely recent loans, because Wells made the choice to give the French pronunciation for those words, and not loans that are more established (e.g. *police*, *Bernadette*, *etiquette* have no French pronunciation in their entry). The stress patterns found in the entries of our dataset were thus coded in binary fashion, final or non-final. We also tested several predictors:

- **Ending:** presence of endings that are known to be islands of regularity for final stress, identified as sequences that occur in at least ten entries;
- **Syntactic categories:** taken from the online *Merriam Webster's* dictionary;

¹⁴⁷ Throughout this section, I will be using the term ‘entries’ to talk about the items studied here, because some of them are clearly borrowed phrases and so calling those items ‘words’ would be misleading.

- **Morphological complexity:** entries that are spaced (e.g. *bon appetit, coup d'état, déjà vu*) or hyphenated (e.g. *beaux-arts, passe-partout, Port-au-Prince*) were treated as complex. It is possible that the orthographic indication of complexity will make those items more likely to be perceived as phrases, and so increase the rate of final primary stress;
- **Weight of the final syllable:** several measures of weight were used, and the one that was retained in the paper because it was the best predictor, was based on mora counts relying mostly on Hammond (1999);
- **Syllable count:** number of phonetic syllables.

Those variables were tested as predictors of the position of primary stress in British and American English, as we were also interested in the differences between those two varieties. An overall look at the data shows a difference between the two varieties, with final primary stress being found in 50% of entries in British English, but in 78% of entries in American English. We ran several statistical analyses to test the effects of those predictors. First, binary logistic regression analyses allowed us to find significant effects of all factors except syntactic category.¹⁴⁸

We found that certain endings were indeed associated to fixed positions of stress. For example, in British English, *-aire* or *-eur/euse* are associated with final primary stress, while *-é(e)* or *-in* are associated with non-final primary stress. We also found that final primary stress was more likely in entries:

- that are complex;
- that have a smaller number of syllables;
- that have heavier final syllables.

In order to analyse those effects more closely, we ran additional analyses using conditional inference trees on the subsets of words which do not contain any of the endings that were found to be associated with a fixed position of stress. The outcomes of those analyses are shown in Figure 41 for British English and in Figure 42 for American English.

¹⁴⁸ Two classes of words were put aside because they were problematic: entries ending with /əw/ (/ow/ in American English), which has been argued to be sometimes monomoraic word-finally, and nasal vowels, whose weight is difficult to establish: should they be treated as underlying short nasal vowels or as short vowels and a nasal consonant? However, the stress patterns found in those words were found to be entirely in line with the rest of the data.

SUBLEXICA IN ENGLISH

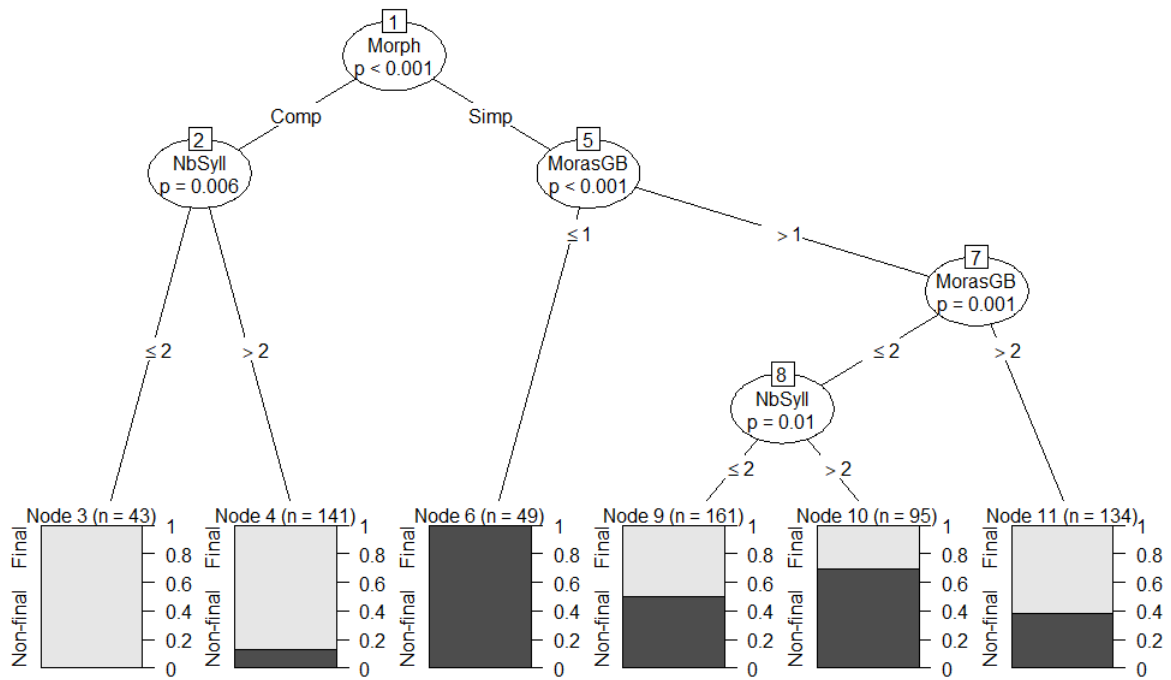


Figure 41. Conditional inference tree for the position of primary stress in French loanwords depending on syllable count, complexity and weight of the final syllable (in moras) in BrE

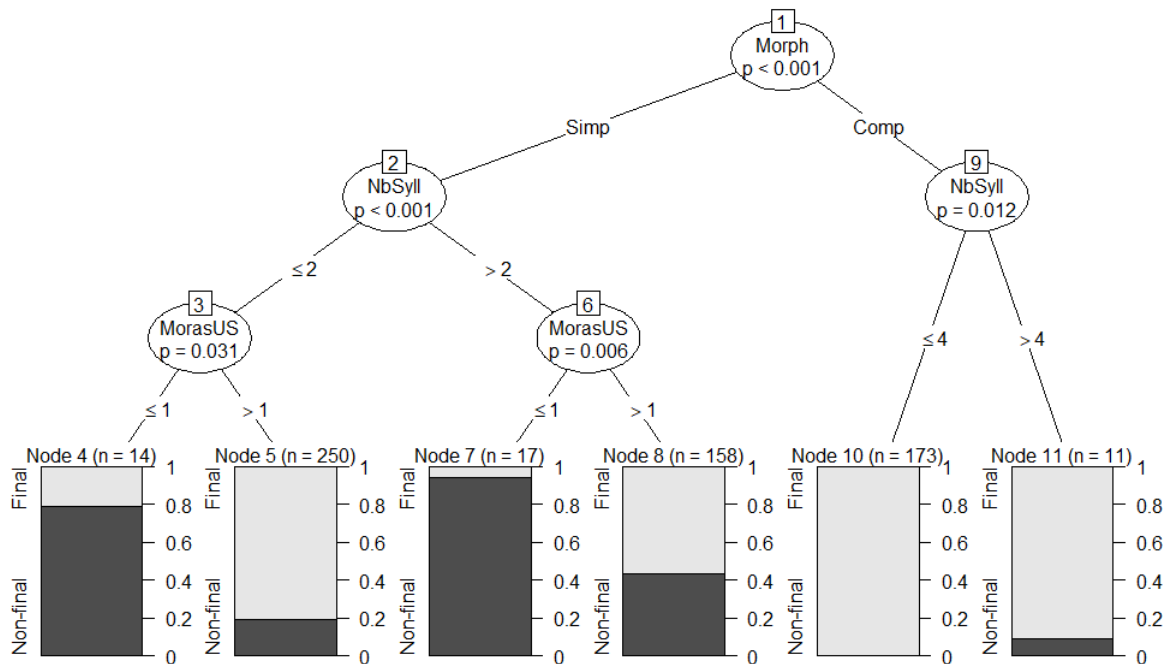


Figure 42. Conditional inference tree for the position of primary stress in French loanwords depending on syllable count, complexity and weight of the final syllable (in moras) in AmE

In both figures, it can be seen that the most significant split in the data with regards the position of primary stress is complexity. Complex entries near-categorically have final stress, with only a few exceptions in entries that are longer. Entries with a final monomoraic syllable

categorically (in British English) or near-categorically (in American English) have non-final primary stress, and there is more final stress in disyllables than in longer entries. Part of the effects of weight that we observe may be seen as problematic. Indeed, final monomoraic syllables end with a short vowel that some would analyse as reduced (e.g. /ə/, /θ/), and so it could be argued that they are monomoraic because they are unstressed and reduced, and so it would be circular to attribute the absence of stress to their monomoraicity as they would most likely not be monomoraic if they were stressed.¹⁴⁹ However, those effects still hold in heavier syllables, as there is more final stress in trimoraic syllables than in bimoraic ones.

Thus, we have found some of the significant predictors of final stress in French loanwords, but there is still a considerable number of entries with non-final primary stress (either as the only pronunciation or as a variant pronunciation), and we may wonder at the stress patterns found in those entries. We found that complex entries predominantly have penultimate stress, and that this could probably be related to the fact that many of those entries have a final disyllabic element with initial stress (e.g. *Crò-Mágnon*, *prèt-à-pórter*, *Quài d'Órsay*). Simple entries predominantly have antepenultimate primary stress when they do not have final primary stress (e.g. *Ármagnàc*, *Báudelàire*, *éntremèts*). We analyse this as Strong Retraction (Lieberman & Prince 1977): a disyllabic foot is built leftwards of the final foot, regardless of the weight of the penult. Although the final syllable does not always have final primary stress, it is still quite often stressed, as was seen in §2.2.3.

Therefore, French loanwords may not always have final primary stress, but they very often have final stress, primary or secondary. The effects of syllable weight and syllable count appear to be gradient and so we suggest that those effects should be captured using a gradient model such as MaxEnt. In the paper, we suggest that when those words constitute §FRENCH in English, this sublexicon has a phonological grammar which differs from §CORE in that NONFINALITY must be weaker and ALIGN-HEAD must be stronger to account for the higher rate of final primary stress in those words than in the rest of the vocabulary. Moreover, an additional constraint requiring the alignment of the right edge of the phonological word with the right edge of a minimal monosyllabic foot may be necessary to account for the fact that final syllables are still generally stressed, even when they are not primary stressed.

Future research may consider additional variables such as frequency, the date of entry in the lexicon, semantic connotations or analogy with existing English words (e.g. is the antepenultimate stress of *pàr éxcellènce* attributable to the stress of the English word *éxcellence*?). It may also be interesting to study the same items using other types of data, as the dictionary data shows a considerable amount of variation, and so it would be interesting to see how they are realized in more naturalistic speech.

¹⁴⁹ One piece of evidence in that direction is the case of final /i/ in Wells (2008), which was analysed as monomoraic. In cases in which an entry with final /i/ has a variant with final stress, the vowel is /i:/, which was analysed as bimoraic (e.g. *gaucherie* /'gəʊʃəri/ ~ /gəʊʃə'ri:/). However, if we assume that final /i/ is unstressed /ij/, then the difference is only one of stress and not necessarily one of mora count.

5.3.2. *Final nuclei and sublexica*

Recently, in the context of the ERSaF project, we decided to reanalyse data from Moore-Cantwell (2020) along with Marie Gabillet. Our reanalysis was presented at the 2024 RFP conference and will be submitted as a squib.

In her paper, Moore-Cantwell presents analyses of primary stress placement in words that are longer than two syllables using the Carnegie Mellon University pronouncing dictionary¹⁵⁰ (henceforth CMU; Weide 1994) and experimental data. In both types of data, she shows that words ending which have /ɨj/ as the nucleus of their final syllable (which she transcribes as /i/) have more antepenultimate primary stress than those which have other nuclei. As her CMU data is available online and as she gave us access to the data from her second experiment, we wanted to test the idea that the asymmetry in stress assignment between different final nuclei is attributable to asymmetries in their distribution across different sublexica. More specifically, we explore the possibility that words in /ɨj/ have more antepenultimate stress because they are found mostly in §CORE, a sublexicon in which antepenultimate stress is more common for words longer than two syllables.

5.3.2.1. *The CMU data*

In her analysis of the CMU data, Moore-Cantwell sorts words into morphologically simple and complex words. She classified the data morphologically automatically using spelling cues and a list of affixes from Teschner & Whitley (2004). She tests the goodness of the categorisation on a sample of 400 words and finds that about 14% of words are miscategorized. We thought that this was potentially a problem to have over a tenth of the data that is morphologically miscategorized, especially considering that certain classes are quite restricted and so this misclassification could seriously affect the results in those classes. Therefore, we manually recoded the data for morphology using the online *OED* and recategorized 564 (7%) of the data:

- 103 words from complex to simplex (e.g. *acacia*, *coriander*, *Uganda*);
- 461 from simplex to complex (e.g. *yesterday*, *nitrogen*, *remarry*, *remembrance*).

Moreover, in the paper, Moore-Cantwell excludes 26 items with secondary stressed /ɨ/ in the last syllable (e.g. *'atmo,sphere* *'daffo,dil*, *'mando,lin*). However, the CMU contains other nuclei that may carry secondary stress in the final syllable, and so we excluded them too. Examples are shown in (116).

(116)	1 syllabic /ɨ/	<i>catapult</i>
	5 syllabic /ɨ/	<i>taciturn, introvert, amateur, extrovert, universe</i>
	45 /ɨj/	<i>quarantine, manatee, Portuguese, magazine...</i>

Then, we classified the simplex words into sublexica relying on the orthographic cues given in Dabouis & P. Fournier (2022). We identified only 19 words as §FRENCH using specific

¹⁵⁰ The CMU is a dictionary of American English lexical items and contains about 134,000 phonetically transcribed entries, with primary and secondary stress annotations. The original paper used a cleaned-up version which excludes very low-frequency items and entries with inflectional morphology.

French endings (*-ie, -ette, -esque*) and spelling-to-sound correspondences (e.g. <qu> - /k/, <ou> - /ʊw/, <ch> - /ʃ/, <ie> - /ij/, <gn> - /nj/). Those words were left out in the rest of the analysis as they were too few. They are listed in (117).

- (117) *brasserie, camaraderie, charlatan, chauvinist, chicanery, chivalry, coterie, couturier, etiquette, insouciant, machination, machinery, mannequin, menagerie, monsignor, picturesque, reverie, rotisserie, tourniquet*

The remaining words were sorted into the other three sublexica using the cues shown in Table 26, along with examples and the number of words thus identified for each sublexicon.

	§FOREIGN	§LEARNED	§CORE
Endings	-Ca, -Va (V≠<e>), -Vh, -i, -o	-um, -is, -im, -in (chemical), -ea	All remaining words
Spelling-to-sound	Non-silent <e#>	<ch> - /k/, <ph> - /f/ <rh> - /ɹ/	
Examples	<i>ballerina, chianti, hyena, nirvana, Pakistani, ukulele</i>	<i>addendum, calculus, emphasis, phobia</i>	<i>agony, dynasty, horizon, mackerel, similar</i>
Number of words categorized in sublexicon	212 (40%)	82 (15%)	236 (45%)

Table 26. Criteria used for the coding of sublexica in the CMU data

We also coded the weight of the penultimate syllable using the categories found in Moore-Cantwell's paper:

- V: open syllable with a short vowel;
- VV: long vowel, possibly in a closed syllable;
- TC: short vowel, optionally followed by L, and an obstruent coda;
- LC: short vowel followed by a sonorant in the coda;
- SC is a short vowel followed by /s/ and another consonant (forming a legal word-initial onset).

We also tested an additional variable, the number of final coda consonants, as it was found to have an effect on stress in long pseudowords in Domahs *et al.* (2014).

We first compared our cleaned-up data with that presented by Moore-Cantwell, who often presents simplex and complex words together. In our data, some categories are not numerous enough to draw any conclusions, but there are three categories (shown in Figure 43) for which we can compare Moore-Cantwell's data to ours.

SUBLEXICA IN ENGLISH

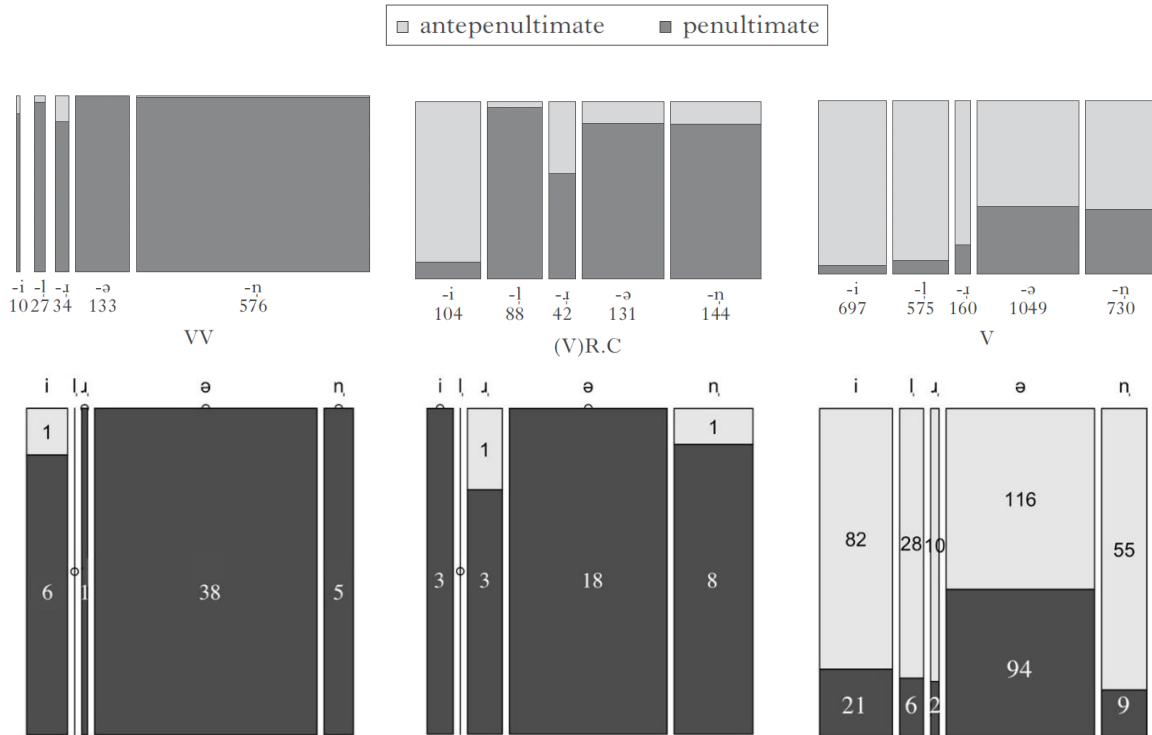


Figure 43. Comparison of CMU data in Moore-Cantwell (top) with our recoded data (bottom)

We can see in Figure 43 that the items ending in /ɪj/ (shown as “i”) do not seem to be really different from other nuclei. If anything, the main difference seems to be the words in /ə/, which seem to have more penultimate stress in all three categories (especially the words in V), an observation that could not be made in Moore-Cantwell’s results, possibly because it was hidden by the inclusion of morphologically complex words in the counts.

Then, we ran several types of regression models and conditional inference tree analyses, with penultimate weight, final nucleus, sublexicon and coda length tested as predictors. Coda length was never found to significantly improve the models and was thus left out. The final nucleus was only found to improve the model if a binary coding was used, contrasting a final /ɪj/ to the other possible nuclei. In the conditional inference tree analyses, we do find a small effect of the final nucleus, which can be seen in Figure 44.

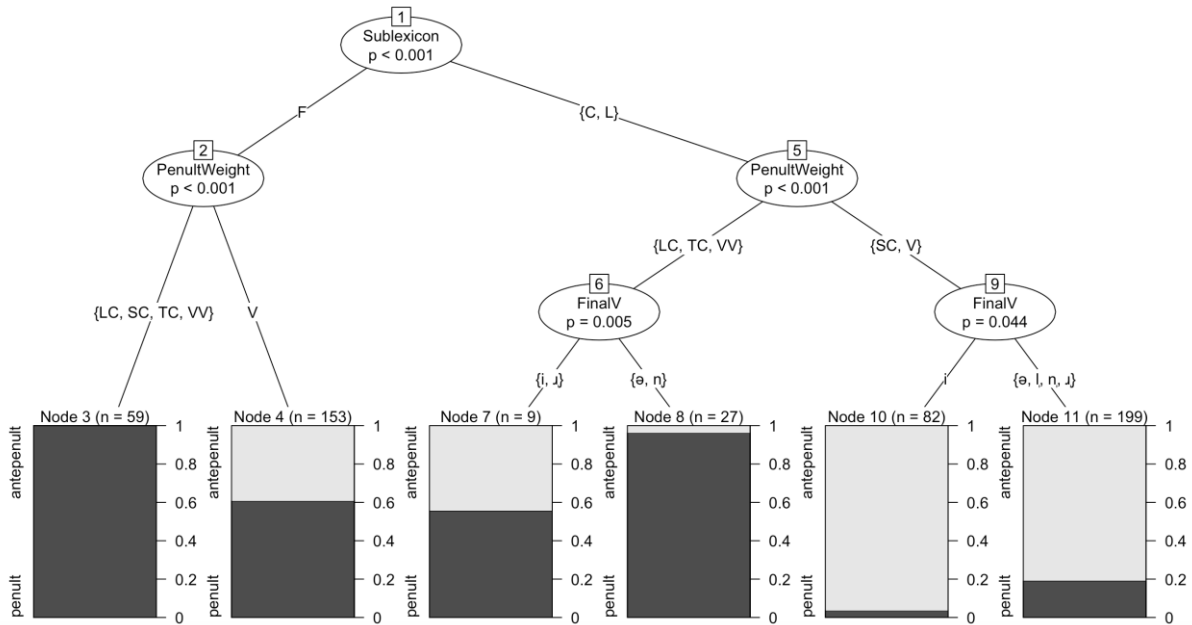


Figure 44. Conditional inference tree analysis of the CMU data

It can be seen in Figure 44 that the first significant split in the data is the variable encoding the different sublexica, and it distinguishes §FOREIGN from the other two sublexica.¹⁵¹ Second, we can see that the second split is penult weight in both subsets of the data, the only difference being the classification of items containing /sC/ clusters. This is not particularly surprising given the well-known issues with the syllabification of such clusters. Finally, we only find a small effect of the final nucleus, which is weakly significant in §CORE and §LEARNED words with light penults and in the subset constituted of the 36 words of those sublexica with heavy penults.

Thus, we do see a small effect of final nuclei in our reanalysis of the CMU data, but it is much more restricted than in Moore-Cantwell’s analysis, and is less important than the effect of sublexica. Now the question is to determine what Dabouis & P. Fournier’s model predicts for the experimental data. First, one possibility is that participants will perceive the pseudowords as foreign and so they may diverge from the stress distributions in the full lexicon to be closer to those found in §FOREIGN, i.e. **they might produce more penultimate stress**. Second, they may seek to match the stress distributions of the lexicon, but also **match the distribution of nuclei among different sublexica**. More specifically, in the experiment, items end either in /ə/ or in /ɪj/ and, as can be seen in Figure 45, those two nuclei have drastically different distributions: /ə/ is almost systematically found in §FOREIGN words while this is only the case of about a third of words in /ɪj/, the rest being almost only words of §CORE.

¹⁵¹ The grouping between §CORE and §LEARNED may have to do with the fact that there are relatively few items in §LEARNED (which is possibly a consequence of the fact that the dataset that is used excludes low-frequency words).

SUBLEXICA IN ENGLISH

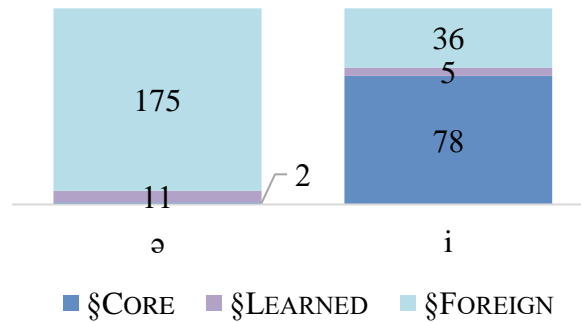


Figure 45. The distribution of sublexica in words ending in /ə/ and /i/ (with no final coda) in the CMU data

Therefore, the prediction could be that participants are more likely to assume that an /ə/-final pseudoword is a §FOREIGN word and thus to assign it penultimate stress than an /ɪ/-final word. Finally, we might expect that the experimental results should be closer to our recorded CMU data if participants make their stress decisions on the basis of simplex words.

5.3.2.2. *The experimental data*

Moore-Cantwell conducted two pseudoword experiments to test the role of final nuclei. In both experiments, the pseudowords were trisyllables made to resemble possible English words and participants were asked to produce those words by concatenating the three syllables that were presented in isolation.¹⁵² They were then asked to choose the option that resembled the most what they had produced out of two possible options, one with penultimate stress and one with antepenultimate stress (e.g. [pæ] [læ] [kə] → ['pæləkə] or [pə'lækə]). The pseudowords were created with light syllables only and, in Experiment 2, penultimate syllables were made heavy by adding a coda (e.g. [pæ] [læ] [kə] / [ki] → [pæ] [læz] [kə] / [ki]).¹⁵³

As can be seen in Figure 46, the distributions found in our recorded CMU are very close to those found in Experiment 2.

¹⁵² This design is modelled after Guion *et al.* (2003) to avoid possible biases from orthography.

¹⁵³ Moore-Cantwell did not control for the legality of the consonant clusters generated as word-medial clusters. It is possible that some of them only occur in morphologically complex words and may have biased participants to treat the words that contain them as such, which could have affected their stress decisions. We ran statistical analyses on the data by adding a variable coding whether or not the cluster is present in simplex words (e.g. the /db/ cluster is found in 157 words in Wells (2008), all of which are morphologically complex: *bedbug*, *bloodbath*, *cardboard*). However, we found no effect of that variable on the position of stress.

SUBLEXICA IN ENGLISH

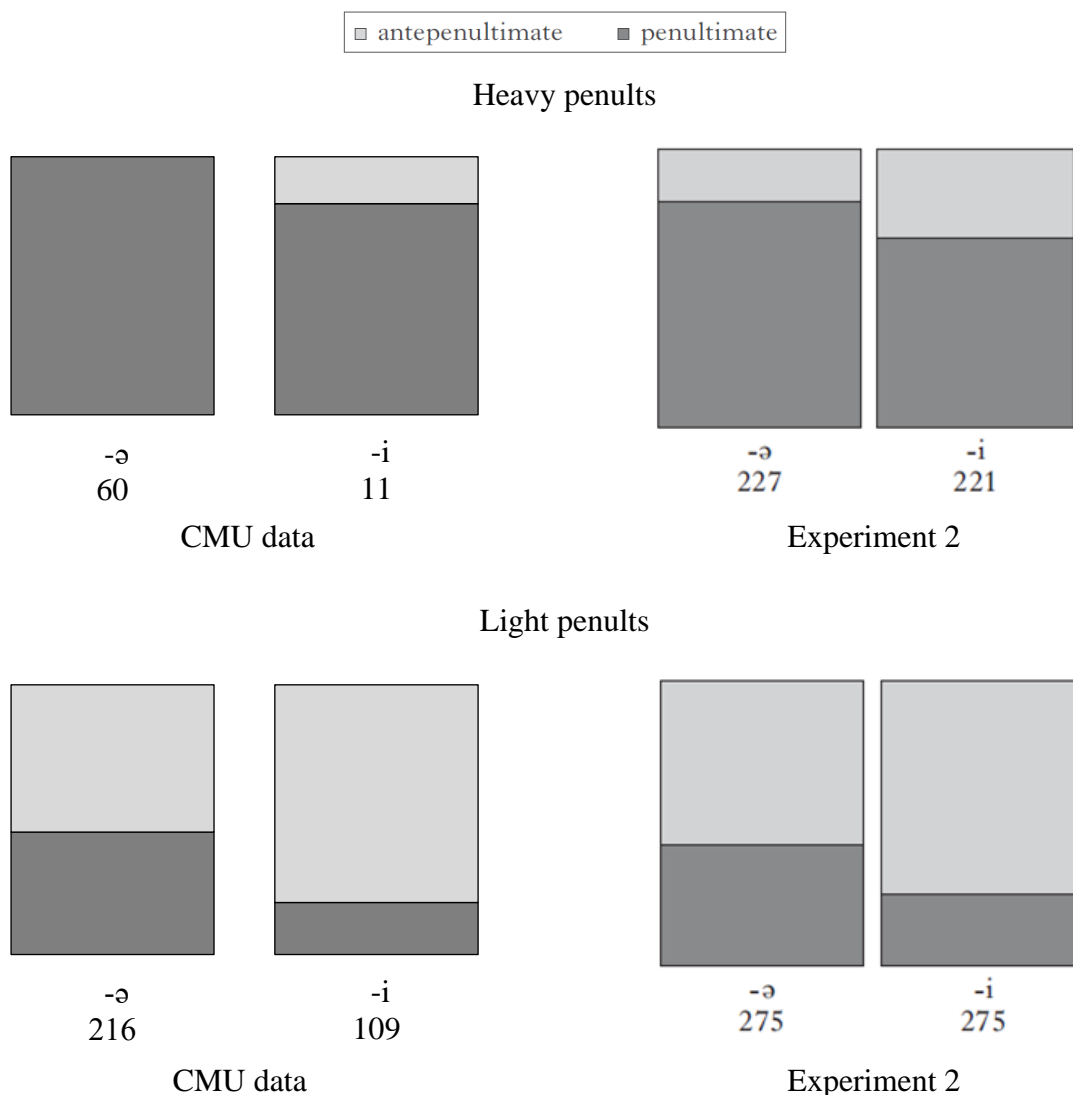


Figure 46. The distribution of the data in the CMU data and in Moore-Cantwell's (2020) Experiment 2

We performed chi-square tests to compare the two distributions. The results of those tests are shown in Table 27, and they show that the only case in which there is a significant difference is the case of words ending in /ə/ with a heavy penult, with more antepenultimate stress being found in Experiment 2 than in the CMU data.

	Heavy penult	Light penult
/i/	Fisher's test ¹⁵⁴ p-value = 0.7283	$\chi^2 = 0.86378$, df = 1, p-value = 0.3527
/ə/	$\chi^2 = 11.923$, df = 1, p-value < .001	$\chi^2 = 0.49983$, df = 1, p-value = 0.4796

Table 27. Statistical tests comparing the distribution of the data in the CMU and in the results from Experiment 2

¹⁵⁴ There were too few values in this category in the CMU data to perform a regular χ^2 test.

This is different from Moore-Cantwell's observations, as she had found that there was more penultimate stress in the results of Experiment 2 than in her treatment of the CMU data. However, we still find more antepenultimate stress for items in /ə/ and a heavy penult, which is something that we have to account for.

Those results show that our first prediction is wrong: we do not find more penultimate stress in the experimental data than in the CMU data. This could be taken to imply that participants did not treat the pseudowords as §FOREIGN, at least not to a greater extent than what is found in the lexicon.¹⁵⁵ The greater proportion of penultimate stress for /ə/-final pseudowords as compared to /ɪj/-final pseudowords supports our second prediction, in which we suggested that participants were more likely to interpret /ə/-final pseudowords as §FOREIGN than /ɪj/-final pseudowords to be consistent with the distribution found in the lexicon. Thus, the difference between /ə/ and /ɪj/ may not have to do with vowel-specific idiosyncrasies as suggested by Moore-Cantwell, but rather with the fact that those vowels have different distributions among sublexica, and that participants matched those distributions, which in turn influenced their stress decisions.

Therefore, our results provide only limited evidence for the role of final nuclei when sublexica are taken into account, and we have put forward an alternative explanation for the difference in stress assignment observed in Moore-Cantwell's experimental data between /ə/-final and /ɪj/-final pseudowords. However, the experiment was not designed to study the role of sublexica in stress assignment. As I have suggested previously, it would be interesting to conduct such experiments, for example by providing definitions along with the pseudowords so as to favour the classification of those pseudowords into a given sublexicon (e.g. a definition referring to a foreign object or tradition might favour the association to §FOREIGN). An indication that this could indeed affect the stress choices made by participants is provided by a replication of Moore-Cantwell's study that she conducted to test a variety of other factors (Moore-Cantwell 2023).¹⁵⁶ In this replication, the pseudowords were presented as American city names, and she found significantly more penultimate stress than in her 2020 study. Many American place names have characteristics that would place them in §FOREIGN¹⁵⁷ and so it is possible that the experimental design favoured the perception of those words as belonging to §FOREIGN which, in turn yielded more responses with penultimate stress than when no specific instruction was given.

¹⁵⁵ However, it is possible that they still treated more experimental items as §FOREIGN than what is found in **their** lexicon, which may not contain as many loanwords as the whole CMU.

¹⁵⁶ This study has not been published anywhere yet and was presented to us in the context of a meeting of the team of the ERSaF project.

¹⁵⁷ There are many place names that are clearly native and have the characteristics described in §3.3.2 (e.g. *Arlington, Clinton, Franklin, Washington*). If we restrict our search to place names that end in vowels, we get names such as *Colorado, Louisiana* or *Missouri*, where we can see final vowels that are unusual in §CORE and the foreign spelling-to-sound correspondences <a>-/ɑ:/, <i>-/ɪj/, <ou>-/uw/ or /ə/ before /i/.

CHAPTER 6

Conclusion

Preparing this document has been the occasion to look back on a 13-year career as a researcher and how I have changed. I have always been working on several topics at the same time, but the diversity of the topics that I have been interested in has increased over time, from the core preoccupations of the Guierrian School – stress and spelling-to-sound correspondences – to diverse issues related to prosodic structure, syllable weight or the organization of the lexicon. The constant characteristics of my work have always been the two introduced at the start of this document – the importance of data and of using various theoretical frameworks – even if their concrete expression has changed over time. Without having pointed out explicitly those characteristics for each topic covered here, I believe that they can clearly be seen throughout this document.

I have covered various topics surrounding the phonology of English words, and when reflecting about some common aspects that may group some of them together, I found two. First, I have been interested in **representations**: the nature of stress and accent, of how we should represent the phonological behaviour of suffixes, knowledge about words made up of semantically impoverished formatives or that of orthography and its relationship with phonology. Second, a significant part of this document has to do with **interfaces**: more than half deals with the interface between morphosyntax and phonology, we have also seen how phonology interacts with orthography, and how the organization of the lexicon into different sublexica affects the phonology of words. In the years to come, I intend to continue exploring some of those topics. I would like to look for more solid evidence of the segmental correlates of stress or the segmental manifestations of different types of suffixes. Within the context of the ERSaF project, we are going to continue exploring the properties of words which contain semantically impoverished formatives, the role of sublexica, and how computational models such as AML or NDL (Naïve Discriminative Learning) can capture their properties. I would also like to explore in more detail the predictions made by the model that we proposed with Pierre Fournier on sublexica. For certain topics, notably semantically impoverished formatives and sublexica, I am particularly interested in looking at individual differences: how do vocabulary size, morphological sensitivity or other sociolinguistic factors affect the phonological patterns that language users produce? Certainly, other topics of interest will come up in the years to come.

As mentioned in the introduction, over the years, the tools and types of data that I have used have changed, from descriptive to inferential statistics, from dictionary-based datasets to experimental and acoustic data. I intend to keep pushing in those directions, by exploring inferential statistics further, and possibly get acquainted with Bayesian statistics, which I suspect might become a new standard in the years to come. I wish to explore computational models such as AML and NDL in more detail and what we can learn from them. For certain processes, it will be necessary to use spoken data and so I will need to learn more about acoustic

CONCLUSION

analysis. Some of the processes that I have investigated using dictionary data should certainly be investigated again using spoken data. I believe that this would be particularly interesting for vowel reduction and pretonic secondary stress.

Finally, I started out in the Guierrian framework and early on learned about some generative models such as *SPE*'s model, Metrical Phonology or Lexical Phonology. The range of theories that I have now become acquainted with has considerably extended, among phonological theories, but also morphological theories or psycholinguistic models on reading or morphological processing. This has provided me with a series of possible 'lenses' through which I can apprehend different processes, different tools to think about the issues that I am interested in, models which generate new issues to think about and predict phenomena that I would not have thought about. I have no intention to fully embrace any model in particular in the near future, but rather to continue exploring what different models tell us about English phonology and what English phonology can bring them.

REFERENCES

- ABASQ, V. (2009). Pre- et post- en anglais contemporain : sémantisme et comportement accentuel. *Corela*, (7–1).
- ABASQ, V. DABOUI, Q. FOURNIER, J.-M. FOURNIER, P. GIRARD, I. MARTIN, M. & VANHOUTTE, S. (2012). Stress in Prefixed Disyllabic Verb/Noun Pairs. Poster Presented at the 20th Manchester Phonology Meeting at the University of Manchester, United Kingdom (25-27 May).
- ABASQ, V. DABOUI, Q. FOURNIER, J.-M. GIRARD, I. & VANHOUTTE, S. (2012). Multicategorical Prefixed Words Stress Behaviour: Variation and Frequency. Paper presented at the 8th International PAC conference (Phonology of Contemporary English) at the University of Toulouse II-Le Mirail, France (29 February-2 March).
- ABERCROMBIE, D. (1976). Stress and Some Other Terms. *Work in Progress*, (9), 51–53.
- ALBER, B. (2020). Word Stress in Germanic. In M. T. Putnam & B. R. Page (Eds.), *The Cambridge Handbook of Germanic Linguistics* (pp. 73–96). Cambridge: CUP.
- ALBRIGHT, A. (2008). Flapometry and palatography: An argument for surface identity between derived forms? Paper Presented at the 4th WOTM Conference (21 June).
- ALBRIGHT, A. (2011). Paradigms. In M. Van Oostendorp, C. J. Ewen, & E. Hume (Eds.), *The Blackwell Companion to Phonology (vol. 4: Phonological Interfaces)* (pp. 1972–2001). Malden, MA: Wiley-Blackwell.
- ANDERSON, S. R. (1981). Why phonology isn't "natural." *Linguistic Inquiry*, (12), 493–539.
- ANDERSON, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- ARNDT-LAPPE, S. (2011). Towards an exemplar-based model of stress in English noun–noun compounds. *Journal of Linguistics*, 47(3), 549–585.
- ARNDT-LAPPE, S. (2015). Word-formation and analogy. In P. O. Müller, I. Ohnheiser, S. Olsen, & F. Rainer (Eds.), *Word-Formation - An International Handbook of the Languages of Europe. Volume 2* (pp. 822–841). Berlin: de Gruyter Mouton.
- ARNDT-LAPPE, S. & BEN HEDIA, S. (2019). Stress in the Family: Reconsidering Stress Preservation in English –ory Adjectives. Poster presented at the 12th Mediterranean Morphology Meeting at the University of Ljubljana, Slovenia (27-30 June).
- ARNDT-LAPPE, S. & DABOUI, Q. (submitted). Secondary stress and morphological structure: New evidence from dictionary and speech data.
- ARNDT-LAPPE, S. & DABOUI, Q. (EDS.). (2018). Secondary stress and morphological structure. Paper presented at the 18th International Morphology Meeting at the Hungarian Academy of Science in Budapest, Hungary (10-13 May).
- ARNDT-LAPPE, S. & ERNESTUS, M. (2020). Morpho-phonological alternations: The role of lexical storage. In V. Pirelli, I. Plag, & W. U. Dressler (Eds.), *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon* (pp. 191–227). Berlin & New York: de Gruyter Mouton.
- ARONOFF, M. (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.
- ARONOFF, M. (1994). *Morphology by Itself: Stems and Inflectional Classes. Linguistic Inquiry Monographs* (Vol. 22).
- ARONOFF, M. & SIMS, A. D. (2023). The relational nature of morphology. In D. Crepaldi (Ed.), *Linguistic Morphology in the Mind and Brain* (pp. 7–25). Abingdon & New York: Routledge.
- ARONOFF, M. & SRIDHAR, S. (1983). Morphological levels in English and Kannada, or, Atarizing Reagan. In J. F. Richardson, M. Marks, & A. Chukerman (Eds.), *Papers from the Parasession on Phonology, Morphology and Syntax. Chicago 22.-23. April 1983* (pp. 3–16).

- BAAYEN, H. R. MILIN, P. FILIPOVIĆ DURDEVIĆ, D. HENDRIX, P. & MARELLI, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118, 438–482.
- BAAYEN, H. R. & SCHREUDER, R. (1999). War and peace: Morphemes and full forms in a noninteractive activation parallel dual-route model. *Brain and Language*, 68, 27–32.
- BAAYEN, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. *Processing*, 2(3), 353.
- BACKLEY, P. (2011). *An Introduction to Element Theory*. Edinburgh: Edinburgh University Press.
- BAESKOW, H. (2006). A revival of romance roots. *Morphology*, 16(1), 3–36.
- BAKER, R. G. & SMITH, P. T. (1976). A psycholinguistic study of english stress assignment rules. *Language and Speech*, 19(1).
- BALLIER, N. (2016). Du dictionnaire lexico-phonétisé aux corpus oraux, quelques problèmes épistémologiques pour l'école de Guierre. *Histoire Épistémologie Langage*, 2(38), 23–40.
- BALLIER, N. & MARTIN, P. (2010). Corrélats prosodiques et acoustiques de la syllabification : le cas du français et de l'anglais. Paper presented at the 8th meeting of the French Phonology Network (RFP) at the University of Orléans, France.
- BASSETTI, B. (2023). *Effects of Orthography on Second Language Phonology: Learning, Awareness, Perception and Production*. Abingdon & New York: Routledge.
- BASSETTI, B. ESCUDERO, P. & HAYES-HARB, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics*, (36), 1–6.
- BAUER, L. LIEBER, R. & PLAG, I. (2013). *The Oxford reference guide to English Morphology*. Oxford: Oxford University Press.
- BELL, A. BRENIER, J. M. GREGORY, M. GIRAND, C. & JURAFSKY, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, (60), 92–111.
- BELL, M. & PLAG, I. (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics*, 48, 485–520.
- BELL, M. & PLAG, I. (2013). Informativity and analogy in English compound stress. *Word Structure*, 6(2), 129–155.
- BEN HEDIA, S. (2019). *Gemination and degemination in English affixation. Studies in laboratory phonology*. Berlin, Germany: Language Science Press.
- BEN HEDIA, S. & PLAG, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics*, 62, 34–49.
- BENUA, L. (1997). *Transderivational Identity: Phonological Relations between Words*. Ph.D. dissertation, University of Massachusetts.
- BERMÚDEZ-OTERO, R. (to appear). /aɪ/-raising, Stratal Phonology, and the life cycle. In P. Honeybone & M. Krämer (Eds.), *Theoretical Phonology and the Phonology of English*. Cambridge: Cambridge University Press.
- BERMÚDEZ-OTERO, R. (1999). Constraint Interaction in Language Change: quantity in English and Germanic. Ph.D. dissertation, University of Manchester.
- BERMÚDEZ-OTERO, R. (2003). The acquisition of phonological opacity. In J. Spenader, A. Eriksson, & Ö. Dahl (Eds.), *Variation within Optimality Theory: Proceedings of the Stockholm Workshop on 'Variation within Optimality Theory'* (pp. 25–36). Stockholm: Department of Linguistics, Stockholm University.
- BERMÚDEZ-OTERO, R. (2004). Raising and Flapping in Canadian English: grammar and acquisition. Paper presented at the CASTL Colloquium at the University of Tromsø, Norway (2 November).
- BERMÚDEZ-OTERO, R. (2007a). On the Nature of the Cycle. Paper presented at the 15th Manchester Phonology Meeting at the University of Manchester, United Kingdom (25 May).

- BERMÚDEZ-OTERO, R. (2007b). Word-final prevocalic consonants in English: representation vs derivation. Paper Presented at the 4th Old World Conference in Phonology at the University of the Aegean in Rhodes, Greece (18-21 January).
- BERMÚDEZ-OTERO, R. (2011). Cyclicity. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology (vol. 4: Phonological Interfaces)* (pp. 2019–2048). Malden, MA: Wiley-Blackwell.
- BERMÚDEZ-OTERO, R. (2012). The Architecture of Grammar and the Division of Labour in Exponence. In J. Trommer (Ed.), *The Phonology and Morphology of Exponence - the State of the Art* (pp. 8–83). Oxford: OUP.
- BERMÚDEZ-OTERO, R. (2013). The Stem-Level Syndrome. Paper Presented to the UPenn Linguistics Department, Speaker Series, Philadelphia (11 April).
- BERMÚDEZ-OTERO, R. (2015). The Monongahela argument. Class Notes for the Course LELA30441 English Phonology Past and Present, University of Manchester (26 November).
- BERMÚDEZ-OTERO, R. (2016). Paradigmatic dependencies without cyclic containment as UR acquisition. Paper presented at the third meeting of Phonological Theory Agora at the University of Tours, France (14-15 October).
- BERMÚDEZ-OTERO, R. (2018a). Diagnostics of the moraic trochee from Proto-Germanic to present-day English. Paper presented at the Workshop on the Foot in the Phonological History of English, ICEHL XX at the University of Edinburgh, United Kingdom (27 August).
- BERMÚDEZ-OTERO, R. (2018b). Stratal Phonology. In S. J. Hannahs & A. R. K. Bosch (Eds.), *The Routledge Handbook of Phonological Theory* (pp. 100–134). Abingdon: Routledge.
- BERMÚDEZ-OTERO, R. (2024). Nonuniformity in Stratal Phonology. Paper Presented at the 21st Old World Conference in Phonology at the University of Leipzig, Germany (14-16 February).
- BERMUDEZ-OTERO, R. & LUIS, A. R. (2009). Cyclic domains and prosodic spans in the phonology of European Portuguese functional morphs. Paper presented at the 6th Old World Conference in Phonology at the University of Edinburgh, United Kingdom (22-24 January).
- BERMUDEZ-OTERO, R. & MCMAHON, A. (2006). English phonology and morphology. In B. Aarts & A. M. S. McMahon (Eds.), *The Handbook of English Linguistics* (pp. 382–410). Oxford: Blackwell.
- BEYERSMANN, E. & GRAINGER, J. (2023). The Role of Embedded Words and Morphemes in Reading. In D. Crepaldi (Ed.), *Linguistic Morphology in the Mind and Brain* (pp. 26–49). Abingdon & New York: Routledge.
- BLOOMFIELD, L. (1933). *Language*. New York: Holt, Rinehart & Winston.
- BOERSMA, P. (2007). Some listener-oriented accounts of h-aspiré in French. *Lingua*, 117(12), 1989–2054.
- BOERSMA, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. In A. Benz & J. Mattausch (Eds.), *Bidirectional Optimality Theory* (pp. 33–72). Amsterdam: Benjamins.
- BOERSMA, P. & WEENIK, D. (2024). PRAAT: Doing Phonetics by Computer [Computer program]. Version 6.4.19, retrieved 4 September 2024 from <http://www.praat.org/>
- BOLINGER, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford: Stanford University Press.
- BOOIJ, G. E. (1977). *Dutch morphology: A study of word formation in generative grammar*. Dordrecht: Foris.
- BOOIJ, G. E. (1983). Principles and parameters in Prosodic Phonology. *Linguistics*, 21, 249–280.

- BOOIJ, G. E. (2010). *Construction Morphology*. Oxford: Oxford University Press.
- BOOIJ, G. E. (2012). Construction Morphology, a brief introduction. *Morphology*, 22(3), 343–346.
- BOOIJ, G. E. & AUDRING, J. (2017). Construction Morphology and the Parallel Architecture of Grammar. *Cognitive Science*, 41 Suppl 2, 277–302.
- BOOIJ, G. E. & RUBACH, J. (1984). Morphological and Prosodic domains in Lexical Phonology. *Phonology Yearbook*, 1, 1–27.
- BOROWSKY, T. (1993). On the word level. In S. Hargus & E. M. Kaisse (Eds.), *Studies in Lexical Phonology* (pp. 199–234). San Diego: Academic Press.
- BOURCIER, G. (1978). *L'orthographe de l'anglais*. PUF.
- BOURCIEZ, E. & BOURCIEZ, J. (1967). *Phonetique française: etude historique*. Paris: Klincksieck.
- BRAM, B. (2011). *Major Total Conversion in English: The Question of Directionality*. Victoria University of Wellington.
- BREISS, C. (2021). *Lexical Conservatism in phonology: theory, experiments, and computational modeling*. Ph.D. dissertation, University of California.
- BREISS, C. (2024). When Bases compete: a voting model of Lexical Conservatism. *Language*, 100(2), 308–358.
- BÜRKI, A. SPINELLI, E. & GARETH GASKELL, M. (2012). A written word is worth a thousand spoken words: The influence of spelling on spoken-word production. *Journal of Memory and Language*, 67(4), 449–467.
- BURZIO, L. (1994). *Principles of English Stress*. New York: Cambridge University Press.
- BURZIO, L. (1996). Surface Constraints Versus Underlying Representation. *Current Trends in Phonology: Models and Methods*, 97–122.
- BUTTERWORTH, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production (Vol. 2)* (pp. 257–294). London: Academic Press.
- BYE, P. & DE LACY, P. (2008). Metrical influences on fortition and lenition. In *Lenition and Fortition* (pp. 173–206).
- CARAMAZZA, A. LAUDANNA, A. & ROMANI, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28(3), 297–332.
- CARNEY, E. (1994). *A survey of English spelling*. London: Routledge.
- CASTANIER, J. (2016). *L'évolution accentuelle du lexique anglais contemporain appréhendée à travers les dictionnaires de prononciation (XVIIIe - XXIe siècles)*. Ph.D. dissertation, Université de Poitiers.
- CASTANIER, J. (2023). Stress Placement in Etymologically Prefixed Disyllabic Noun–Verb Pairs Revisited: A Semantic and Diachronic Approach. In E. Yamada, N. Ballier, J.-M. Fournier, & A. Przewozny (Eds.), *New Perspectives on English Word Stress* (pp. 83–116). Edinburgh: Edinburgh University Press.
- CHADELAT, J.-M. (2000). *Valeur et fonctions des mots français en anglais à l'époque contemporaine*. Paris, Montréal: L'Harmattan.
- CHEN, C.-Y. (2014). Direction of stress-shift in noun-verb pairs and progressions in American and British English. *English Linguistics*, 2(31), 401–438.
- CHEN, C.-Y. (2017). Word frequency, entry date and entry status in relation to stress-shifts in English noun-verb pairs. *English Linguistics*, 2(33), 231–277.
- CHEN, M. (1970). Vowel Length Variation as a Function of the Voicing of the Consonant Environment. *Phonetica*, 22(3), 129–159.
- CHEVROT, J.-P. (1998). Acquisition phonologique tardive et apprentissage de la lecture-écriture. *Travaux de l'Institut de Phonétique de Strasbourg*, (28), 19–37.
- CHOMSKY, N. (2000). Minimalist inquiries: The framework: Step by step. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Essays on Minimalist Syntax in Honor of Howard Lasnik*

- (pp. 89–155). Cambridge, MA: MIT Press.
- CHOMSKY, N. & HALLE, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.
- CHOMSKY, N. HALLE, M. & LUKOFF, F. (1956). On accent and juncture in English. In M. Halle, H. G. Lunt, H. McLean, & C. H. van Schooneveld (Eds.), *For Roman Jakobson: Essays on the occasion of his sixtieth birthday, 11 October 1956* (pp. 65–80). The Hague: Mouton.
- CHUNG, S. (1983). Transderivational Constraints in Chamorro Phonology. *Language*, 59(1), 35–66.
- CLOPPER, C. G. & TURNBULL, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Nieburh, B. Schuppler, & M. Zellers (Eds.), *Rethinking Reduction: Interdisciplinary Perspectives on Conditions, Mechanisms, and Domains for Phonetic Variation* (pp. 25–72). Berlin, New York: de Gruyter Mouton.
- COCH, D. BARES, J. & LANDERS, A. (2013). ERPs and morphological processing: The N400 and semantic composition. *Cognitive, Affective and Behavioral Neuroscience*, 13(2), 355–370.
- COCH, D. HUA, J. & LANDERS-NELSON, A. (2020). All morphemes are not the same: accuracy and response times in a lexical decision task differentiate types of morphemes. *Journal of Research in Reading*, 43(3), 329–346.
- COLLIE, S. (2007). *English Stress Preservation and Stratal Optimality Theory*. Ph.D. dissertation, University of Edinburgh.
- COLLIE, S. (2008). English Stress Preservation: the Case for ‘Fake Cyclicity.’ *English Language and Linguistics*, 12(03), 505–532.
- CORETTA, S. (2020). *Vowel duration and consonant voicing: A production study*. Ph.D. dissertation, University of Manchester.
- CREEMERS, A. (2023). Morphological Processing in Spoken-Word Recognition. In D. Crepaldi (Ed.), *Linguistic Morphology in the Mind and Brain* (pp. 50–64). Abingdon & New York: Routledge.
- CRUTTENDEN, A. (2014). *Gimson’s Pronunciation of English* (8th edition). Oxon & New York: Routledge.
- DABOUIS, Q. (2012). *Accentuation et contrainte en anglais contemporain : le cas des adjectifs suffixés en -al*. Master’s thesis, Université de Tours.
- DABOUIS, Q. (2014). English Stress and Underlying Representations. *Proceedings of the First Postgraduate and Academic Researchers in Linguistics at York*, (1), 1–15.
- DABOUIS, Q. (2016a). Is the Adjectival Suffix -al a Strong Suffix ? *Anglophonia* [Online], (21).
- DABOUIS, Q. (2016b). *L’accent secondaire en anglais britannique contemporain*. Ph.D. dissertation, University of Tours.
- DABOUIS, Q. (2016c). Syllable Weight and Secondary Stress in English Suffixal Derivatives. In Q. Dabouis, A. Ndione, & C. Ternisien (Eds.), *Proceedings of the LLL Young Researchers Conference* (pp. 6–23).
- DABOUIS, Q. (2017). Semantically Opaque Prefixes and English Phonology. *Paper Presented at the 14th Old World Conference in Phonology on 20-22th February at the University of Düsseldorf*.
- DABOUIS, Q. (2018). The pronunciation of vowels with secondary stress. *Corela* [Online], 16(2).
- DABOUIS, Q. (2019). When Accent Preservation Leads to Clash. *English Language and Linguistics*, 23(2), 363–404.
- DABOUIS, Q. (2020a). Catégorie syntaxique et accentuation en anglais : le cas des dissyllabes verbe-nom. Paper Presented at the LRL’s Seminar, Online (17 April).
- DABOUIS, Q. (2020b). Secondary stress in contemporary British English: An overview. *Anglophonia* [Online], (30).
- DABOUIS, Q. (2022a). Failles (in)visibles ? Morphologie opaque et phonologie en anglais. *Paper*

- Presented at the Workshop of the ALAES and ALOES during the 61th Congress of the SAES, at the Université Clermont Auvergne, Clermont-Ferrand, France (2-4 June).
- DABOUIS, Q. (2022b). John T. Jensen (2022) *The Lexical and Metrical Phonology of English: The Legacy of The Sound Pattern of English*. Cambridge, New York: Cambridge University Press. Pp. xv + 379. *Phonology*, 39(4), 725–734.
- DABOUIS, Q. (2023a). English Phonology and the Literate Speaker: Some Implications for Lexical Stress. In N. Ballier, J.-M. Fournier, A. Przewozny, & E. Yamada (Eds.), *New Perspectives on English Word Stress* (pp. 117–153).
- DABOUIS, Q. (2023b). Morphological gradience in phonology. Paper Presented at the 30th Manchester Phonology Meeting at the University of Manchester, United Kingdom (25-27 May).
- DABOUIS, Q. (2023c). Opaque yet still complex: Proper names in English. Paper Presented at the 20th OCP Conference (Old World Conference in Phonology) at the University of Tours, France (25-27 January).
- DABOUIS, Q. (2023d). Phonological evidence for morphological complexity in English proper names. *Anglophonia* [Online], (36).
- DABOUIS, Q. ENGUEHARD, G. FOURNIER, J.-M. & LAMPITELLI, N. (2020). The English “Arab Rule” without feet. *Acta Linguistica Academica*, 1(67), 121–134.
- DABOUIS, Q. ENGUEHARD, G. & LAMPITELLI, N. (2019). *Ternary rhythm as a complex morphological domain*. Paper presented at the RecPhon2019 conference at the Autonomous University of Barcelona, Spain (21-22 November).
- DABOUIS, Q. & FOURNIER, J.-M. (to appear). An empirical study of vowel reduction and preservation in British English. *Phonology*.
- DABOUIS, Q. & FOURNIER, J.-M. (2023). The Stress Patterns of English verbs: Syllable Weight or Morphology? In N. Ballier, J.-M. Fournier, A. Przewozny, & E. Yamada (Eds.), *New Perspectives on English Word Stress* (pp. 154–191). Edinburgh: Edinburgh University Press.
- DABOUIS, Q. & FOURNIER, J.-M. (2024a). Opaque morphology and phonology: Historical prefixes in English. FirstView article, *Journal of Linguistics*.
- DABOUIS, Q. FOURNIER, J.-M. FOURNIER, P. & MARTIN, M. (2023). English Word Stress and the Guierrian School. In N. Ballier, J.-M. Fournier, A. Przewozny, & E. Yamada (Eds.), *New Perspectives on English Word Stress* (pp. 53–82). Edinburgh: Edinburgh University Press.
- DABOUIS, Q. FOURNIER, J.-M. & GIRARD, I. (2017). Ternarity is not an Issue: Secondary Stress is Left Edge Marking. Paper presented at the Fringe Meeting of the 25th Manchester Phonology Meeting – Phonological Theory Agora Dataset Workshop at the University of Manchester, United Kingdom (24 May).
- DABOUIS, Q. FOURNIER, J.-M. & MARTIN, M. (2014). Testing Parameters for Stress Placement : The Case of Dissyllabic Prefixed Verb / Noun Pairs. Paper presented at the 11th Old World Conference in Phonology at the University of Leiden, Netherlands (22-25 January).
- DABOUIS, Q. & FOURNIER, P. (2022). English PhonologiES? In V. Arigne & C. Rocq-Migette (Eds.), *Modèles et modélisation en linguistique / Models and Modelisation in Linguistics* (pp. 215–258). Bruxelles, Belgique: Peter Lang.
- DABOUIS, Q. & FOURNIER, P. (2024b). Stress in French Loanwords in British and American English. FirstView article, *Journal of Linguistics*.
- DABOUIS, Q. & GABILLET, M. (2024). On stress and accent in English. Paper Presented at the 21st ALOES Conference at the Université Sorbonne Paris Nord (29-30 March).
- DABOUIS, Q. GLAIN, O. & NAVARRO, S. (submitted). Gemination and degemination of /r/ in British and American English.
- DABOUIS, Q. & MARTIN, M. (2022). La classification des suffixes: de Guierre à aujourd’hui.

- Paper presented at the workshop "40 Ans de Guierre", online (31 March).
- DAHAK, A. (2006). Quel statut pour un corpus ‘ oral ’ écrit ? Le dictionnaire de prononciation comme corpus (quasi-) exhaustif de la langue . In *Actes des IXèmes rencontres Jeunes Chercheurs de l’Ecole Doctorale 268 “Langage et Langues”* (pp. 1–5). Paris: Université de la Sorbonne Nouvelle.
- DAHAK, A. (2011). *Etude diachronique, phonologique et morphologique des syllabes inaccentuées en anglais contemporain*. Ph.D. dissertation, Université de Paris Diderot.
- DAVIS, M. (2011). Corpus of American Soap Operas. URL: <https://www.english-corpora.org/soap/>
- DAVIS, S. (to appear). A Foot-Based Approach to English Schwa Syncope. In P. Honeybone & M. Krämer (Eds.), *Theoretical Phonology and the Phonology of English*. Cambridge: Cambridge University Press.
- DAVIS, S. (2005). Capitalistic v. Militaristic: The Paradigm Uniformity Effect Reconsidered. In L. Downing, A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory* (pp. 106–121). Oxford: Oxford University Press.
- DAVIS, S. & CHO, M.-H. (2003). The distribution of aspirated stops and /h/ in American English and Korean: an alignment approach with typological implications. *Linguistics*, 41(4), 607–652.
- DAWSON, N. RASTLE, K. & RICKETTS, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning Memory and Cognition*, 44(4), 645–654.
- DERWING, B. L. (1992). Orthographic Aspects of Linguistic Competence. In *The Linguistics of Literacy* (pp. 193–210). Amsterdam - Philadelphia: John Benjamins Publishing.
- DESCHAMPS, A. (1994). *De l’écrit à l’oral et de l’oral à l’écrit*. Paris: Ophrys.
- DESCHAMPS, A. & O’NEIL, M. (2007). Obituary: Lionel Guierre. *Language Sciences*, (29), 492–495.
- DI SCIULLO, A.-M. & WILLIAMS, E. (1987). *On the Definition of the Word*. Cambridge, MA: MIT Press.
- DICKERSON, W. B. (1980). Bisyllabic laxing rule: vowel prediction in linguistics and language learning. *Language Learning*, 30, 317–329.
- DIEPENDAELE, K. SANDRA, D. & GRAINGER, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition*, 6(37), 895–908.
- DIXON, R. M. W. (1977a). *A Grammar of Yidiny*. Cambridge: Cambridge University Press.
- DIXON, R. M. W. (1977b). Some Phonological Rules in Yidiny. *Linguistic Inquiry*, 8(1), 1–34.
- DOMAHS, U. PLAG, I. & CARROLL, R. (2014). Word Stress Assignment in German, English and Dutch: Quantity-Sensitivity and Extrametricality Revisited. *The Journal of Comparative Germanic Linguistics*, 17(1), 59–96.
- DOWTY, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67(3), 547–619.
- DUANMU, S. (2010). Onset and the Weight-Stress Principle in English. In *Linguistic essays in honor of Professor Tsu-Lin Mei on his 80th birthday* (pp. 1–31).
- DURAND, J. NAVARRO, S. & VIOLLAIN, C. (2015). ‘R’-sandhi in English: how to constrain theoretical approaches. *Global Communication Studies, World Englishes*, 2, 103–132.
- EHRI, L. C. & WILCE, L. S. (1980). The Influence of Orthography on Reader’s Conceptualization of the Phonemic Structure of Words. *Applied Psycholinguistics*, 2, 371–385.
- EMBICK, D. CREEMERS, A. & GOODWIN DAVIES, A. (2022). Morphology and The Mental Lexicon: Three questions about decomposition. In A. Papafragou, J. C. Trueswell, & L. R. Gleitman (Eds.), *The Oxford Handbook of the Mental Lexicon* (pp. 77–97). Oxford: Oxford University Press.

- ENGEMANN, M. & PLAG, I. (2021). Phonetic reduction and paradigm uniformity effects in spontaneous speech. *The Mental Lexicon*, 16(1).
- ERNESTUS, M. & NEIJT, A. (2008). Word length and the location of primary word stress in Dutch, German, and English. *Linguistics*, 46(3), 507–540.
- FADEL, Y. (2023). *Stress placement in English disyllabic verbs and nouns*. Master's thesis, Université Clermont Auvergne.
- FARNETANI, E. TORSELLO, C. T. & COSI, P. (1988). English compound versus non-compound noun phrases in discourse: An acoustic and perceptual study. *Language and Speech*, (31), 157–180.
- FAUST, N. & TORRES-TAMARIT, F. (2017). Stress and final /n/ deletion in Catalan: Combining Strict CV and OT. *Glossa*, 2(1).
- FELDMAN, LAURIE BETH O'CONNOR, P. A. & DEL PRADO MARTÍN, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, 16(4), 684–691.
- FERRAGNE, E. (2021). *Du groupe à l'individu, du corpus à l'expérimentation, du spectrogramme au deep learning pour la phonétique*. Habilitation à diriger des recherches, Aix-Marseille Université.
- FIDELHOLTZ, J. (1966). Vowel reduction in English. Manuscript.
- FIDELHOLTZ, J. (1975). Word Frequency and Vowel Reduction in English. *Chicago Linguistic Society*, 11, 200–213.
- FLYNN, D. (to appear). Setting internal boundaries in English words. In P. Honeybone & M. Krämer (Eds.), *Theoretical Phonology and the Phonology of English*. Cambridge: Cambridge University Press.
- FODOR, J. (1983). *The Modularity of the Mind*. Cambridge, MA: MIT-Bradford.
- FORSTER, K. I. & AZUMA, T. (2000). Masked priming for prefixed words with bound stems: Does submit prime permit? *Language and Cognitive Processes*, 15(4–5), 539–561.
- FOURNIER, J.-M. (1992). Un modèle de lecture de la voyelle accentuée. *Actes de l'atelier de Linguistique Du 32e Congrès de La SAES (Strasbourg), Saint-Etienne : CIEREC*.
- FOURNIER, J.-M. (1994). La prononciation de la voyelle accentuée /ɔ/ en anglais RP. In *7ème Colloque d'Avril sur l'anglais oral* (pp. 151–169). Villetaneuse: Université de Paris-Nord, CELDA, diffusion APLV.
- FOURNIER, J.-M. (1996). La reconnaissance morphologique. In *8ème Colloque d'Avril sur l'anglais oral* (pp. 45–75). Villetaneuse: Université de Paris-Nord, CELDA, diffusion APLV.
- FOURNIER, J.-M. (1998). Que contraignent les terminaisons contraignantes? In *Topiques, Nouvelles recherches en linguistique anglaise, Travaux XCIII du CIEREC* (pp. 45–75). Université de Saint-Etienne.
- FOURNIER, J.-M. (2007). From a Latin syllable-driven stress system to a Romance versus Germanic morphology-driven dynamics: in honour of Lionel Guierre. *Language Sciences*, 29, 218–236.
- FOURNIER, J.-M. (2010a). Accentuation lexicale et poids syllabique en anglais: l'analyse erronée de Chomsky et Halle. Paper Presented at the 8th Meeting of the French Phonology Network (RFP) at the University of Orléans, France (1–3 July).
- FOURNIER, J.-M. (2010b). *Manuel d'anglais oral*. Paris: Ophrys.
- FOURNIER, J.-M. (2024). A la recherche du système perdu. Paper Presented at the “Phonologie En Système” Workshop at the University of Tours (13 June).
- FOURNIER, P. (2011). *Accentuation et prononciation des suffixés en -ous en anglais contemporain*. Ph.D. dissertation, Université de Tours.
- FOURNIER, P. (2016). Nouvelles perspectives sur l'accentuation des emprunts en anglais

- contemporain. *Itinéraires, 2015–2*.
- FOURNIER, P. (2018). Stress Assignment in Italian Loanwords in English and its Impact on the Stressing of Foreign Words by Native English Speakers. *Corela HS-24 Multicultural Spoken English*.
- FOURNIER, P. (2021a). Les emprunts au gaélique écossais en anglais. In S. Juillet-Garzon, P. Fournier, & A. Fiasson (Eds.), *L'Écosse : la différence* (pp. 215–229). Presses Universitaires de Franche-Comté.
- FOURNIER, P. (2021b). Transcription et formalisation de l'accentuation des mots « italiens » en anglais contemporain. In F. Toupin & C. Collin (Eds.), *Transcrire, écrire, formaliser 2* (pp. 53–68). Presses Universitaires de Rennes.
- FOURNIER, P. (2024). The impact of source languages on the stressing of loanwords in English. *Lexis: Journal of English Lexicology*, (23).
- FOX, A. (2000). *Prosodic Features and Prosodic Structure*. New York: Oxford University Press.
- FRAUENFELDER, U. H. & SCHREUDER, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. E. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 165–185). Dordrecht: Kluwer Academic.
- FRITCHE, R. SHATTUCK-HUFNAGEL, S. & SONG, J. Y. (2021). Do adults produce phonetic variants of /t/ less often in speech to children? *Journal of Phonetics*, 87.
- FUDGE, E. (1984). *English Word Stress*. London: G. Allen & Unwin.
- GAFFIOT, F. (1934). *Dictionnaire illustré Latin-Français*. Paris: Hachette.
- GARCIA, G. D. & GOAD, H. (to appear). Weight effects and the role of the foot: English vs. Portuguese. *Laboratory Phonology*.
- GICK, B. & WILSON, I. (2001). Pre-Liquid excrescent schwa: What happens when vocalic targets conflict. In *EUROSPEECH 2001 - SCANDINAVIA - 7th European Conference on Speech Communication and Technology* (pp. 3–7).
- GIEGERICH, H. J. (1992). The Limits of Phonological Derivation: Spelling Pronunciations and Schwa in English. *Linguistische Berichte*, 142, 413–436.
- GIEGERICH, H. J. (1999). *Lexical Strata in English: Morphological Causes, Phonological Effects*. Cambridge: Cambridge University Press.
- GIEGERICH, H. J. (2009). The English compound stress myth. *Word Structure*, 2(1), 1–24.
- GOAD, H. (2012). SC Clusters are (almost always) coda-initial. In *Linguistic Review* (Vol. 29).
- GOLDWATER, S. & JOHNSON, M. (2003). Learning OT Constraint Rankings Using a Maximum Entropy Model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, (1997), 111–120.
- GONET, W. & STADNICKA, L. (2005). Vowel Clipping in English. *Speech and Language Technology*, 8, 77–86.
- GRAEBER, D. & WENGROW, D. (2021). *The dawn of everything: A new history of humanity*. New York: Farrar, Straus and Giroux.
- GREEN, A. (2007). Phonology Limited. *Linguistics in Potsdam*, (27).
- GREENWELL, B. M. MCCARTHY, A. J. BOEHMKE, B. C. & LIU, D. (2018). Residuals and diagnostics for binary and ordinal regression models: An introduction to the sure package. *R Journal*, 10(1).
- GUIERRE, L. (1979). *Essai sur l'accentuation en anglais contemporain : Éléments pour une synthèse*. Ph.D. dissertation, Université Paris-VII.
- GUIERRE, L. (1983). L'accent préfère-t-il les longues ? *Méthodes En Linguistique Anglaise, Travaux N° XXXIX*. Saint-Etienne: CIEREC.
- GUIERRE, L. (1990). Mots composés anglais et agrégats consonantiques. In *5ème Colloque d'Avril sur l'anglais oral* (pp. 59–72). Villetaneuse: Université de Paris-Nord, CELDA,

diffusion APLV.

- GUIERRE, L. (2000). Pourquoi la morpho-phonologie. *Publications de l'Université de Pau*, 32–46.
- GUION, S. G. CLARK, J. J. HARADA, T. & WAYLAND, R. P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, 46(4), 403–427.
- GUSSENHOVEN, C. (1991). The English Rhythm Rule as an accent deletion rule. *Phonology*, 8, 1–35.
- GUSSENHOVEN, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- GUSSENHOVEN, C. (2011). Sentential Prominence in English. In M. Van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology (vol. 2: Suprasegmental and Prosodic Phonology)*2 (pp. 2778–2806). Malden, MA: Wiley-Blackwell.
- GUYOT-TALBOT, A. (2003). *Les préfixes et la reconnaissance morphologique*. Ph.D. dissertation, Université Paris VII.
- HALLE, M. & KENSTOWICZ, M. (1991). The free element condition and cyclic vs. noncyclic stress. *Linguistic Inquiry*, 22, 457–501.
- HALLE, M. & KEYSER, S. (1971). *English Stress: Its Form, Its Growth, and its Role in Verse*. New York: Harper & Row.
- HALLE, M. & MARANTZ, A. (1993). Distributed Morphology and the Pieces of Inflection. *The View from Building 20: Essays in Honor of Sylvain Bromberger*, 111–176.
- HALLE, M. & VERGNAUD, J.-R. (1978). *Metrical Structures in Phonology*. MIT.
- HALLE, M. & VERGNAUD, J.-R. (1987). *An Essay on Stress*. Cambridge, MA: MIT.
- HAMANN, S. (2020). One phonotactic restriction for speaking, listening and reading. In M. Evertz-Rittich & F. Kirchhoff (Eds.), *Geschriebene und gesprochene Sprache als Modalitäten eines Sprachsystems* (pp. 57–78). Berlin & Boston: Walter de Gruyter.
- HAMANN, S. & COLOMBO, I. E. (2017). A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. *Natural Language & Linguistic Theory*, (35), 683–714.
- HAMMOND, M. (1989). Cyclic Secondary Stresses in English. In *Proceedings of the annual meeting of the West Coast Conference on Formal Linguistics* (Vol. 8, pp. 139–153). Stanford, California: Stanford Linguistics Association, Stanford University.
- HAMMOND, M. (1999). *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. (J. Durand, Ed.). Oxford: Oxford University Press.
- HAMMOND, M. (2003). Frequency, cyclicity, and optimality. Paper Presented at the Second International Korean Phonology Conference at Seoul National University, South Korea.
- HAMMOND, M. (2006). Prosodic Phonology. In B. Aarts & A. McMahon (Eds.), *Handbook of English Linguistics* (pp. 411–432). Malden: Blackwell.
- HARRIS, J. (1990). Derived phonological contrasts. In S. Ramsaran (Ed.), *Studies in the pronunciation of English. A commemorative volume in honour of A. C. Gimson* (pp. 89–105). London: Routledge.
- HARRIS, J. (1994). *English Sound Structure*. Oxford: Blackwell.
- HARRIS, J. (2004). Release the captive coda: the foot as a domain of phonetic interpretation. In J. K. Local, R. Ogden, & R. A. M. Temple (Eds.), *Phonetic Interpretation* (pp. 103–129). Cambridge: Cambridge University Press.
- HARRIS, J. (2005). Reduction as information loss. In P. Carr, J. Durand, & C. J. Ewen (Eds.), *Headhood, Elements, Specification and Contrastivity: Phonological papers in honour of John Anderson* (pp. 119–132). Amsterdam: John Benjamins Publishing.
- HARRIS, J. (2013). Wide-domain r -effects in English. *Journal of Linguistics*, 49(2), 329–365.

- HARRIS, Z. S. (1951). *Structural Linguistics*. Chicago, IL: University of Chicago Press.
- HASPELMATH, M. (2002). *Understanding Morphology*. Londres: Arnold.
- HASPELMATH, M. (2020). The morph as a minimal linguistic form. *Morphology*, 30(2), 117–134.
- HASPELMATH, M. (2023). Defining the word. *Word*, 69(3), 283–297.
- HAY, J. (2001). Lexical Frequency in Morphology: Is Everything Relative? *Linguistics: An Interdisciplinary Journal of the Language Sciences*, 39(6), 1041–1070.
- HAY, J. (2003). *Causes and Consequences of Word Structures*. London: Routledge.
- HAY, J. (2007). The phonetics of “un.” In *Lexical Creativity, Texts and Contexts2* (pp. 39–57). Amsterdam - Philadelphia: John Benjamins Publishing.
- HAY, J. & BAAYEN, H. (2003). Phonotactics, Parsing and Productivity. *Italian Journal of Linguistics*, 15(1), 99–130.
- HAY, J. & BAAYEN, H. R. (2002). Parsing and Productivity. In G. E. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2001* (pp. 203–235). Dordrecht: Kluwer.
- HAYES, B. (1980). *A Metrical Theory of Stress Rules*. Ph.D. dissertation, MIT.
- HAYES, B. (1981). *A Metrical Theory of Stress Rules*. Bloomington: Indiana University Linguistics Club.
- HAYES, B. (1982). Extrametricality and English Stress. *Linguistic Inquiry*, 13(2), 227–276.
- HAYES, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: The University of Chicago Press.
- HAYES, B. (2000). Gradient well-formedness in Optimality Theory. In J. Dekkers, F. van der Leeuw, & J. van de Weijer (Eds.), *Optimality Theory: Phonology, Syntax, and Acquisition* (pp. 88–120). Oxford: Oxford University Press.
- HERMENT, S. (2010). The pedagogical implications of variability in transcription, the case of [i] and [u]. In A. Henderson (Ed.), *English Pronunciation : Issues and Practices (EPIP). Proceedings of the first international conference*. (pp. 177–188). Presses universitaires Savoie Mont Blanc.
- HEYER, V. & KORNISHOVA, D. (2018). Semantic transparency affects morphological priming ... eventually. *The Quarterly Journal of Experimental Psychology*, 71(5), 1112–1124.
- HOCKETT, C. F. (1947). Problems of Morphemic Analysis. *Language*, 23(4), 321–343.
- HORNE, M. (1990). Empirical evidence for a deletion formulation of the rhythm rule in English. *Linguistics*, 28(5), 959–982.
- HYMAN, L. (1975). *Phonology: Theory and Analysis*. New York: Holt, Rinehart & Winston.
- ITO, J. & MESTER, A. (1992). Weak layering and word binarity. In T. Honma, M. Okazaki, T. Tabata, & S. Tanaka (Eds.), *A New Century of Phonology and Phonological Theory. A Festschrift for Professor Shosuke Haraguchi on the Occasion of His Sixtieth Birthday* (pp. 26–65). Tokyo: Kaitakusha.
- ITÔ, J. & MESTER, A. (1995). Japanese Phonology. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 816–838). Oxford: Blackwell.
- ITÔ, J. & MESTER, A. (1999). The phonological lexicon. *The Handbook of Japanese Linguistics*, 62–100.
- ITÔ, J. & MESTER, A. (2009). The extended prosodic word. In B. Kabak & J. Grijzenhout (Eds.), *Phonological Domains: Universals and Derivations* (pp. 135–194). Berlin, New York: Mouton de Gruyter.
- JACKENDOFF, R. (1975). Morphological and semantic regularities in the lexicon. *Language*, 51(3), 639–671.
- JACKENDOFF, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- JACKENDOFF, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

- JACKENDOFF, R. (2010). The Ecology of English Noun-Noun Compounds. In *Meaning and the Lexicon* (pp. 413–445). Oxford: Oxford University Press.
- JAEGER, J. J. (1986). On the acquisition of abstract representations for English vowels. *Phonology Yearbook*, 3, 71.
- JARMULOWICZ, L. (2006). School-aged children's phonological production of derived English words. *Journal of Speech, Language, and Hearing Research*, 49, 294–308.
- JARMULOWICZ, L. TARAN, V. L. & HAY, S. E. (2008). Lexical frequency and third-graders' stress accuracy in derived English word production. *Applied Psycholinguistics*, 29(02).
- JENSEN, J. T. (2000). Against ambisyllabicity. *Phonology*, 17(2), 187–235.
- JENSEN, J. T. (2022). *The Lexical and Metrical Phonology of English. The Lexical and Metrical Phonology of English*. Cambridge, New York: Cambridge University Press
- JONES, D. (1963). *English Pronouncing Dictionary* (12th edition). London: Dent.
- JONES, D. (2006). *Cambridge English Pronouncing Dictionary* (17th edition). Cambridge: Cambridge University Press.
- KAGER, R. (1989). *A Metrical Theory of Stress and Destressing in English and Dutch*. Ph.D. dissertation, University of Utrecht.
- KAHN, D. (1976). Syllable-based generalizations in English phonology. Ph.D. dissertation, Indiana University Linguistics Club.
- KAISSE, E. M. & SHAW, P. A. (1985). On the theory of Lexical Phonology. *Phonology Yearbook*, 2, 1–30.
- KATAMBA, F. & STONHAM, J. T. (2006). *Morphology. Modern linguistics series* (2nd ed.). Houndsmills, Basingstoke, Hampshire; New York, N.Y.: Palgrave Macmillan.
- KAYE, A. S. (2005). Gemination in English. *English Today*, 21(2), 43–55.
- KAYE, J. (1995). Derivations and interfaces. *Frontiers of Phonology*, 289–332.
- KELLY, M. H. (2004). Word onset patterns and lexical stress in English. *Journal of Memory and Language*, 50(3), 231–244.
- KELLY, M. H. & BOCK, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 389–403.
- KENSTOWICZ, M. (1996). Base-identity and Uniform Exponence: Alternatives to Cyclicity. In J. Durand & B. Laks (Eds.), *Current Trends in Phonology: Models and Methods. Volume 1* (pp. 363–393). Salford: European Studies Research Institute.
- KENSTOWICZ, M. (1997). Quality-sensitive stress. *Rivis*, 9(1), 157–187.
- KENYON, J. S. & KNOTT, T. A. (1953). *A Pronouncing Dictionary of American English*. Springfield, MA: Merriam.
- KING, H. & FERRAGNE, E. (2020). Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English. *Journal of Phonetics*, 80.
- KIPARSKY, P. (1979). Metrical Structure Assignment is Cyclic. *Linguistic Inquiry*, 10(3), 421–441.
- KIPARSKY, P. (1982). From Cyclic Phonology to Lexical Phonology. In H. Van der Hulst & N. Smith (Eds.), *The Structure of Phonological Representations I* (pp. 131–175). Dordrecht: Foris.
- KIPARSKY, P. (1985). Some consequences of lexical phonology. *Phonology, Yearbook 2*, 85–130.
- KIPARSKY, P. (2000). Opacity and cyclicity. *Linguistic Review*, 17, 351–365.
- KIPARSKY, P. (2005). Paradigm uniformity constraints. Manuscript, Stanford University. Available at www.stanford.edu/~kiparsky/Papers/LexConservatism.pdf
- KIPARSKY, P. (2015). Stratal OT: A synopsis and FAQs. In Y. E. Hsiao & L. H. Wee (Eds.), *Capturing Phonological Shades* (pp. 2–44). Newcastle upon Tyne: Cambridge Scholars Publishing.
- KÖHNLEIN, B. (2015). The morphological structure of complex place names: the case of Dutch.

- Journal of Comparative Germanic Linguistics*, 3(18), 183–212.
- KRASKA-SZLENK, I. (2007). Analogy: the Relation Between Lexicon and Grammar. *LINCOM Studies in Theoretical Linguistics*. Munich: LINCOM Europa.
- KTORI, M. MOUSIKOU, P. & RASTLE, K. (2018). Cues to Stress Assignment in Reading Aloud. *Journal of Experimental Psychology*, 147(1), 36–61.
- KTORI, M. TREE, J. J. MOUSIKOU, P. COLTHEART, M. & RASTLE, K. (2016). Prefixes repel stress in reading aloud: Evidence from surface dyslexia. *Cortex*, 74(October 2015), 191–205.
- KUNTER, G. (2011). *Compound Stress in English - The Phonetics and Phonology of Prosodic Prominence*. *Linguistische Arbeiten* (Vol. 539). Berlin: de Gruyter Mouton.
- KUPERMAN, V. PLUYMAEKERS, M. & ERNESTUS, M. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America*, 121, 2261–2271.
- LABOV, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(2).
- LAKS, B. (2005). La liaison et l'illusion. *Langages*, 158(2), 101–125.
- LAUDANNA, A. BURANI, C. & CERMELE, A. (1994). Prefixes as Processing Units. *Language and Cognitive Processes*, 9(3), 295–316.
- LAVOIE, L. M. & COHN, A. C. (1999). Sesquisyllables of English: the structure of vowel-liquid syllables. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 109–112.
- LEES, R. B. (1961). *The Phonology of Modern Standard Turkish*. Bloomington: Indiana University Press.
- LEVSHINA, N. (2021). Conditional Inference Trees and Random Forests. In Magali Paquot, & Stefan T. Gries (Eds.). *Practical handbook of corpus linguistics*. New York: Springer, 611–643.
- LIBERMAN, M. (1975). *The Intonational System of English*. Ph.D. dissertation, MIT.
- LIBERMAN, M. & PRINCE, A. S. (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry*, 8(2), 249–336.
- LIEBER, R. (2004). *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.
- LINDSEY, G. (2019). *English after RP: Standard British Pronunciation Today*. Cham: Palgrave Macmillan.
- LIU, D. & ZHANG, H. (2018). Residuals and Diagnostics for Ordinal Regression Models: A Surrogate Approach. *Journal of the American Statistical Association*, 113(522).
- LOWENSTAMM, J. (1996). CV as the Only Syllable Type. In J. Durand & B. Laks (Eds.), *Current Trends in Phonology Models and Methods* (pp. 419–442). University of Salford: European Studies Research Institute.
- LOWENSTAMM, J. (1999). The beginning of the Word. In J. R. Rennison & K. Kühnammer (Eds.), *Phonologica 1996: Syllables !?* (pp. 153–166). The Hague: Thesus.
- MARANTZ, A. (2013). No escape from morphemes in morphological processing. *Language and Cognitive Processes*, 28(7), 905–916.
- MARELLI, M. TRAFICANTE, D. & BURANI, C. (2020). Reading morphologically complex words: Experimental evidence and learning models. In V. Pirelli, I. Plag, & W. U. Dressler (Eds.), *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon* (pp. 553–592). Berlin & Boston: de Gruyter Mouton.
- MARSLÉN-WILSON, W. D. BOZIC, M. & RANDALL, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes*, 23(3), 394–421.
- MARSLÉN-WILSON, W. D. KOMISARJEVSKY TYLER, L. WAKSLER, R. & OLDER, L. (1994). Morphology and Meaning in the English Mental Lexicon. *Psychological Review*, 101(1),

- MARTIN, M. (2011). *De l'accentuation lexicale en anglais australien standard contemporain*. Ph.D. dissertation, Université de Tours.
- MASCARÓ, J. (2016). Morphological exceptions to vowel reduction in central catalan and the problem of the missing base. *Catalan Journal of Linguistics*, 15, 27–51.
- MCCARTHY, J. J. (1991). Synchronic Rule Inversion. *Annual Meeting of the Berkeley Linguistics Society*, 17, 192–207.
- MCCARTHY, J. J. & PRINCE, A. S. (1995). Faithfulness and Reduplicative Identity. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *University of Massachusetts Occasional Papers in Linguistics 18* (pp. 249–384). Amherst: GLSA.
- MCCARTHY, J. J. & PRINCE, A. S. (1999). Faithfulness and identity in prosodic morphology. In R. Kager, H. Van der Hulst, & W. Zonneveld (Eds.), *The Prosody-Morphology Interface* (pp. 218–309). Cambridge: Cambridge University Press.
- McKINNON, R. ALLEN, M. & OSTERHOUT, L. (2003). Morphological decomposition involving non-productive morphemes: ERP evidence. *NeuroReport*, 14(6), 883–886.
- MCMAHON, A. (2001). Review of *The Phonology of English: a Prosodic Optimality-Theoretic Approach*. Oxford: Oxford University Press by Michael Hammond (1999). *Phonology*, (18), 421–426.
- MCMAHON, A. (2007). Who's afraid of the vowel shift rule? *Language Sciences*, 29(2–3), 341–359.
- MCWHORTER, J. (2003). *Doing our own thing: the degradation of language and music and why we should, like, care*. New York: Gotham Books.
- MELINGER, A. (2001). *Morphological complexity in English prefixed words: an experimental investigation. Methodology*. Ph.D. dissertation, University of New York at Buffalo.
- MINKOVA, D. (1997). Constraint ranking in Middle English stress-shifting. *English Language and Linguistics* 1, 1(1), 135–175.
- MINKOVA, D. (2006). Old and Middle English Prosody. In A. van Kemenade & B. Los (Eds.), *The Handbook of the History of English* (pp. 95–125). Oxford: Blackwell.
- MOHANAN, K. P. (1982). *Lexical Phonology*. Ph.D. dissertation, MIT.
- MOHANAN, K. P. (1986). *The Theory of Lexical Phonology*. Dordrecht: Reidel.
- MOMPEÁN-GONZALEZ, J. & MOMPEÁN-GUILLAMÓN, P. (2009). /r/-liaison in English. *Cognitive Linguistics*, 20(4), 733–776.
- MONTGOMERY, S. (2001). *The Case for Synchronic Orthographic Primacy: the effect of literacy on phonological processing*. Ph.D. dissertation, University of Edinburgh.
- MONTGOMERY, S. (2005). Lax Vowels, Orthography and /ə/: The Need for Orthographic Primacy. *Linguistische Berichte*, 201, 14–64.
- MOORE-CANTWELL, C. (2016). *The Representation of Probabilistic Phonological Patterns : Neurological , Behavioral , and Computational Evidence From the English Stress System*. Ph.D. dissertation, University of Massachusetts Amherst.
- MOORE-CANTWELL, C. (2020). Weight and final vowels in the English stress system. *Phonology*, 37(4), 657–695.
- MOORE-CANTWELL, C. (2023). Rule stability in wug tests. Paper presented at the meeting of the ERSaF project of the 21 November.
- MORLEY, R. L. & SMITH, B. J. (2023). A Reanalysis of the Voicing Effect in English: With Implications for Featural Specification. *Language and Speech*, 66(4).
- MYERS, J. (1999). Lexical phonology and the lexicon. Manuscript, National Chung Cheng University.
- NAVARRO, S. (2017). *Le /r/ en anglais*. Dijon: Editions universitaires de Dijon.
- NESPOR, M. & VOGEL, I. (1986). *Prosodic Phonology*. Foris: Dordrecht.
- NESSLY, L. (1982). Vowel tenseness in English: Another look. *Language Learning*, 32(2), 393–

- NEUMAN, Y. (2009). *L'influence de l'écriture sur la langue*. Paris III - Sorbonne nouvelle.
- NEWELL, H. (2021). Deriving Level 1/Level 2 affix classes in English: Floating vowels, cyclic syntax. *Acta Linguistica Academica*, 68(1/2), 31–76.
- NEWELL, H. (2023). The too many tools problem in phonology: Multiple Phonologies + Prosodic Hierarchy = unfalsifiable theories. Invited Talk at the 30th Manchester Phonology Meeting at the University of Manchester, United Kingdom (25-27 May).
- NIDA, E. A. (1949). *Morphology: The descriptive analysis of words* (2nd edition). Ann Arbor: University of Michigan Press.
- OH, E. (2013). Effects of boundary strength on geminate duration in English. *Studies in Phonetics, Phonology and Morphology*, 19(3), 457–478.
- OH, G. E. & REDFORD, M. A. (2012). The production and phonetic representation of fake geminates in English. *Journal of Phonetics*, 40(1), 82–91.
- PAGLIANO, C. (2003). *L'épenthèse consonantique en français. Ce que la syntaxe, la sémantique et la morphologie peuvent faire à la phonologie*. Ph.D. dissertation, Université de Nice.
- PASTIZZO, M. J. & FELDMAN, L. B. (2004). Morphological processing: A comparison between free and bound stem facilitation. *Brain and Language*, 90(1–3), 31–39.
- PATER, J. (1994). Against the underlying specification of an exceptional English stress pattern. *Toronto Working Papers in Linguistics*, (13), 95–121.
- PATER, J. (1995). On the Nonuniformity of Weight-to-Stress and Stress Preservation Effects in English. Manuscript, McGill University.
- PATER, J. (2000). Non-uniformity in English Secondary Stress: the Role of Ranked and Lexically Specific Constraints. *Phonology*, 17(2), 237–274.
- PATER, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science*, 33(6), 999–1035.
- PATER, J. (2016). Universal Grammar with Weighted Constraints. In J. J. McCarthy & J. Pater (Eds.), *Harmonic Grammar and Harmonic Serialism* (pp. 1–35). London: Equinox.
- PERRE, L. & ZIEGLER, J. C. (2008). On-line activation of orthography in spoken word recognition. *Brain Research*, 1188, 132–8.
- PICARD, M. (2003). On the emergence and resolution of hiatus. *Folia Linguistica Historica*, 24(1–2), 47–57.
- PIERREHUMBERT, J. B. (2006). The statistical basis of an unnatural alternation. *Laboratory Phonology 8, Varieties of Phonological Competence*, 8, 81–107.
- PINKER, S. (1999). *Words and Rules: The Ingredients of Language*. New York: Harper Collins.
- PLAG, I. (1999). *Morphological Productivity: Structural Constraints in English Derivation*. Berlin: Mouton de Gruyter.
- PLAG, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press.
- PLAG, I. (2006). The Variability of Compound Stress in English: Structural, Semantic and Analogical Factors. *English Language and Linguistics*, 1(10), 143–172.
- PLAG, I. & BEN HEDIA, S. (2018). The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In S. Arndt-Lappe, A. Braun, C. Moulin, & E. Winter-Froemel (Eds.), *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and the Role of Discourse-related Factors* (pp. 93–116). Berlin, New York: Mouton de Gruyter.
- PLAG, I. & KUNTER, G. (2007). The phonetics of primary vs. secondary stress in English. Manuscript.
- PLAG, I. KUNTER, G. & LAPPE, S. (2007). Testing Hypotheses About Compound stress Assignment in English: A Corpus-Based Investigation. *Corpus Linguistics and Linguistic Theory*, 2(3), 199–232.
- PLAG, I. KUNTER, G. LAPPE, S. & BRAUN, M. (2008). The role of semantics, argument structure,

- and lexicalization in compound stress assignment in English. *Language*, 84(4), 760–794.
- PLAG, I. KUNTER, G. & SCHRAMM, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, 39(3), 362–374.
- POLDAUF, I. (1984). *English Word Stress*. Oxford: Pergamon Press.
- POLGÁRDI, K. (2015). Vowels, glides, off-glides and on-glides in English: A Loose CV analysis. *Lingua*, 158, 9–34.
- POPPER, K. (2002). *The Logic of Scientific Discovery*. London & New York: Routledge.
- POSTAL, P. M. (1968). *Aspects of phonological theory*. New York: Harper & Row.
- PRINCE, A. S. (1983). Relating to the grid. *Linguistic Inquiry*, (14), 19–100.
- PRINCE, A. S. (1991). Quantitative consequences of rhythmic organization. *Papers from the Twenty-Sixth Regional Meeting of the Chicago Linguistics Society, Vol. 2: The Parasession on the Syllable in Phonetics and Phonology*, 26(1991), 355–398.
- PRINCE, A. S. & SMOLENSKY, P. (1993). Optimality Theory: Constraint Interaction in Generative Grammar. Manuscript, Rutgers University and University of Colorado.
- R CORE TEAM. (2023). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- RAFFELSIEFEN, R. (1999). Diagnostics for Prosodic Words Revisited: The Case of Historically Prefixed Words in English. In T. A. Hall & U. Kleinhenz (Eds.), *Studies on the Phonological Word (Current Issues in Linguistic Theory 174)* (pp. 133–201). Amsterdam: John Benjamins Publishing.
- RAFFELSIEFEN, R. (2004). Absolute ill-formedness and other morphophonological effects. *Phonology*, 21(1), 91–142.
- RAFFELSIEFEN, R. (2005). Paradigm Uniformity Effects versus Boundary Effects. In L. J. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory* (pp. 211–262). Oxford: Oxford University Press.
- RAFFELSIEFEN, R. (2007). Morphological Word Structure in English and Swedish: the Evidence from Prosody. *Fifth Mediterranean Morphology Meeting*, 209–268.
- RAFFELSIEFEN, R. (2010). Idiosyncrasy, Regularity, and Synonymy in Derivational Morphology: Evidence for Default Word Interpretation Strategies. In S. Olsen (Ed.), *New Impulses in Word-Formation* (pp. 173–232). Hamburg: Buske.
- RAFFELSIEFEN, R. (2015). Phonological Restrictions on English Word-formation. In P. O. Müller, I. Ohnheiser, S. Olsen, & F. Rainer (Eds.), *Word-Formation: An International Handbook of the Languages of Europe* (pp. 894–917). Berlin, Boston: Mouton de Gruyter.
- RAFFELSIEFEN, R. (2023). Morphophonological Asymmetries in Affixation. In P. Ackema, S. Benjallabah, E. Bonet, & A. Fábregas (Eds.), *The Wiley Blackwell Companion to Morphology*. Hoboken, New Jersey: John Wiley & Sons.
- RASTLE, K. (2019). The place of morphology in learning to read in English. *Cortex*, 116, 45–54.
- RASTLE, K. & COLTHEART, M. (2000). Lexical and Nonlexical Print-to-Sound Translation of Disyllabic Words and Nonwords. *Journal of Memory and Language*, 42, 342–364.
- RASTLE, K. & DAVIS, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23(7–8), 942–971.
- RIERA, M. & ROMERO, J. (2007). Schwa in American English V+/I/ Sequences: Speaking Rate Effects. In *Actes des 5èmes Journées d'Etudes Linguistiques (27-28 juin 2007 à l'Université de Nantes)*. *Schwa(s)* (pp. 131–136).
- RIPLEY, B. VENABLES, B. BATES, D. M. HORNIK, K. GEBHARDT, A. & FIRTH, D. (2019). MASS: support functions and datasets for venables and Ripley's MASS. R package version 7.3-51.5. Available from <https://CRAN.R-project.org/package=MASS>. *R Topics Documented*.
- ROSS, J. R. (1972). A Reanalysis of English Word Stress. In M. K. Brame (Ed.), *Contributions to Generative Phonology* (pp. 229–323). Austin & London: University of Texas Press.

- RYAN, K. M. (2014). Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language*, 90(2).
- SCHANE, S. (2007). Understanding English word accentuation. *Language Sciences*, 29(2–3), 372–384.
- SCHEER, T. (2004). *A Lateral Theory of Phonology. Vol 1: What is CVCV, and Why Should It Be ?* Berlin: Mouton de Gruyter.
- SCHEER, T. (2008). Why the Prosodic Hierarchy is a diacritic and why the interface must be direct. In J. Hartmann, V. Hegedüs, & H. van Riemsdijk (Eds.), *Sounds of Silence: Empty Elements in Syntax and Phonology* (pp. 145–192). Amsterdam: Elsevier.
- SCHEER, T. (2009). External sandhi: What the initial CV is initial of. *Studi e Saggi Linguistici* 47, (47), 43–82.
- SCHEER, T. (2011). *A Guide to Morphosyntax-Phonology Interface Theories. How Extra-Phonological Information is Treated in Phonology since Trubetzkoy's Grenzsignale.* Berlin: Mouton de Gruyter.
- SCHEER, T. (2012a). Chunk definition in phonology: prosodic constituency vs. phase structure. In M. Bloch-Trojnar & A. Bloch-Rozmej (Eds.), *Modules and Interfaces* (pp. 221–253). Lublin: Wydawnictwo KUL.
- SCHEER, T. (2012b). *Direct Interface and One-Channel Translation. A Non-Diacritic Theory of the Morphosyntax-Phonology Interface.* Berlin: Mouton de Gruyter.
- SCHEER, T. (2013). The Corpus : A Tool among Others. *CORELA* [Online], HS-13.
- SCHEER, T. (2014). The initial CV: Herald of a non-diacritic interface theory. In S. Bendjaballah, N. Faust, M. Lahrouchi, & N. Lampitelli (Eds.), *The Form of Structure, the Structure of Form. Essays in Honor of Jean Lowenstamm* (pp. 315–330). Amsterdam: Benjamins.
- SCHEER, T. & CYRAN, E. (2018a). Interfaces in Government Phonology. In S. J. Hannahs & A. R. K. Bosch (Eds.), *The Routledge Handbook of Phonological Theory* (pp. 293–324). Abingdon & New York: Routledge.
- SCHEER, T. & CYRAN, E. (2018b). Syllable structure in Government Phonology. In S. J. Hannahs & A. R. K. Bosch (Eds.), *The Routledge Handbook of Phonological Theory* (pp. 262–292). Abingdon & New York: Routledge.
- SCHEER, T. & KULA, N. C. (2018). Government Phonology: Element theory, conceptual issues and introduction. In S. J. Hannahs & A. R. K. Bosch (Eds.), *The Routledge Handbook of Phonological Theory* (pp. 226–261). Abingdon & New York: Routledge.
- SCHEER, T. & SÉGÉRAL, P. (2020). Elastic s+C and Left-moving Yod in the Evolution from Latin to French. *Probus*, 32(2).
- SCHREUDER, R. & BAAYEN, H. R. (1995). Modeling morphological processing. In Laurie B Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 131–154). Hillsdale, New Jersey: Lawrence Erlbaum.
- SÉGÉRAL, P. & SCHEER, T. (2001). La Coda-Miroir. *Bulletin de La Société de Linguistique de Paris*, (96), 107–152.
- SÉGÉRAL, P. & SCHEER, T. (2008). The Coda Mirror , stress and positional parameters. In J. B. de Carvalho, T. Scheer, & P. Ségéral (Eds.), *Lenition and Fortition* (pp. 483–518). Berlin: Mouton de Gruyter.
- SEILER, A. & ARNDT-LAPPE, S. (2024). Predicting stress in English verbs – analogy, morphology, and the level of abstraction of phonological patterns. Paper presented at LabPhon19 at Hanyang University in Seoul, South Korea (June 27-29).
- SELKIRK, E. O. (1980a). Prosodic Domains in Phonology: Sanskrit Revisited. In M. Aronoff & M.-L. Kean (Eds.), *Juncture* (pp. 107–129). Saratoga: Anma Libri.
- SELKIRK, E. O. (1980b). The Role of Prosodic Categories in English Word Stress. *Linguistic Inquiry*, 11(3), 563–605.

- SELKIRK, E. O. (1982). *The Syntax of Words*. Cambridge, MA: MIT Press.
- SELKIRK, E. O. (1984). *Phonology and Syntax: The Relation between Sound and Structure. Current studies in linguistics series*. Cambridge: MIT Press.
- SELKIRK, E. O. (1995). The Prosodic Structure of Function Words. In J. N. Beckman, S. Urbanczyk, & L. Walsh (Eds.), *Papers in Optimality Theory* (pp. 439–469). Amherst: University of Massachusetts Occasional Papers.
- SHERMAN, D. (1975). Noun-Verb Stress Alternation: an Example of the Lexical Diffusion of Sound Change in English. *Linguistics*, 13(159).
- SIEGEL, D. C. (1974). *Topics in English Morphology*. Ph.D. dissertation, MIT.
- SIEGEL, D. C. (1979). *Topics in English Morphology*. New York: Garland.
- SKOUSEN, R. (1989). *Analogical Modeling of Language*. Dordrecht: Kluwer.
- SKOUSEN, R. (2002). An overview of Analogical Modeling. In R. Skousen, D. Lonsdale, & D. B. Parkinson (Eds.), *Analogical Modeling* (pp. 11–26). Amsterdam / Philadelphia: John Benjamins.
- SKOUSEN, R. (2009). Expanding Analogical Modeling into a general theory of language prediction. In J. P. Blevins & J. Blevins (Eds.), *Analogy in Grammar* (pp. 164–184). Oxford: OUP.
- SKOUSEN, R. STANFORD, T. & GLENN, N. (2013). *Algorithm::AM*.
- SMOLENSKY, P. & GOLDRICK, M. (2016). *Gradient symbolic representations in grammar: The case of French liaison*. Johns Hopkins University and Northwestern University, *Rutgers Optimality Archive* 1552. 1–37.
- SONDEREGGER, M. & NIYOGI, P. (2013). Variation and change in English noun/verb pair stress: Data and dynamical systems models. In A. C. L. Yu (Ed.), *Origins of Sound Change* (pp. 262–284). Oxford: Oxford University Press.
- SPENCER, A. (1996). *Phonology: Theory and Description*. Oxford: Blackwell.
- SPROAT, R. & FUJIMURA, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21(3).
- STANTON, J. & STERIADE, D. (2014). Stress windows and Base Faithfulness in English suffixal derivatives. Paper presented at the 22nd Manchester Phonology Meeting at the University of Manchester, United Kingdom (29-31 May).
- STANTON, J. & STERIADE, D. (2021). Markedness drives base selection experimental evidence for the pseudo-cycle. Paper presented at the 28th Manchester Phonology Meeting at the University of Manchester, United Kingdom (26-28 May).
- STEIN, S. D. & PLAG, I. (2020). Farewell to relative frequency? A closer look at frequency effects on the acoustic duration of English derivatives. Paper presented at the Words in the World International Conference 2020, online (16-18 October).
- STERIADE, D. (1997). Lexical Conservatism. *Linguistics in the Morning Calm*, 157–179.
- STERIADE, D. (1999). Lexical conservatism in French adjectival liaison. In B. E. Bullock, J.-M. Authier, & L. A. Reed (Eds.), *Formal Perspectives in Romance Linguistics* (pp. 243–270). Amsterdam: John Benjamins.
- STERIADE, D. & STANTON, J. (2020). Productive pseudo-cyclicity and its significance. Paper presented at LabPhon 17, online (6-8 July).
- STEVENS, P. & PLAUT, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychonomic Bulletin and Review*.
- STEWART, T. W. (2016). *Contemporary Morphological Theories. Contemporary Morphological Theories*. Edinburgh: Edinburgh University Press.
- STOCKWELL, R. & MINKOVA, D. (2001). *English Words: History and Structure*. Cambridge: Cambridge University Press.
- SZIGETVÁRI, P. (to appear). The vowel inventory of British English. In P. Honeybone & M. Krämer (Eds.), *Theoretical Phonology and the Phonology of English*. Cambridge:

- Cambridge University Press.
- SZIGETVÁRI, P. (2007). Branching onsets and syncope in English. *Language Sciences*, 29(2–3), 408–425.
- SZIGETVÁRI, P. (2016). No diphthong, no problem. In J. Szpyra-Kozłowska & E. Cyran (Eds.), *Phonology, its faces and interfaces* (pp. 123–141). Budapest: Department of English Linguistics, Eötvös Loránd University.
- SZIGETVÁRI, P. (2017). English stress is binary and lexical. In G. Lindsey & A. Nevins (Eds.), *Sonic Signatures: Studies dedicated to John Harris* (pp. 263–276). Amsterdam: John Benjamins.
- SZIGETVÁRI, P. (2018). Stressed schwa in English. *The Even Yearbook*, (13), 81–95.
- SZIGETVÁRI, P. (2020). Posttonic stress in English. In K. Jaskula (Ed.), *Phonological and Phonetic Explorations* (pp. 163–189). Lublin: Wydawnictwo KUL.
- SZPYRA, J. (1989). *The Morphology-Phonology Interface: Cycles, Levels and Words*. London: Routledge.
- TAFT, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, 7(4), 263–272.
- TAFT, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, 9(3), 271–294.
- TAFT, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology Section A*, 57(4), 745–765.
- TAFT, M. & FORSTER, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 638–647.
- TAFT, M. & HAMBLY, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, (24), 320–335.
- TAFT, M. HAMBLY, G. & KINOSHITA, S. (1986). Visual and auditory recognition of prefixed words. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, (38A), 351–366.
- TAFT, M. & KOUGIOUS, P. (2004). The processing of morpheme-like units in monomorphemic words. *Brain and Language*, 90(1–3), 9–16.
- TANNER, J. SONDEREGGER, M. STUART-SMITH, J. & DATA CONSORTIUM, S. (2019). Vowel duration and the voicing effect across English dialects. *Toronto Working Papers in Linguistics*, 41(1).
- TANNER, J. SONDEREGGER, M. STUART-SMITH, J. & FRUEHWALD, J. (2020). Toward “English” Phonetics: Variability in the Pre-consonantal Voicing Effect Across English Dialects and Speakers. *Frontiers in Artificial Intelligence*, 3.
- TESCHNER, R. V. & WHITLEY, M. S. (2004). *Pronouncing English: a stress-based approach with CD-ROM*. Washington, D.C.: Georgetown University Press.
- TOKAR, A. (2018). Secondary stress variation in contemporary English. *AAA - Arbeiten Aus Anglistik Und Amerikanistik*, 43(2), 15–47.
- TOKAR, A. (2019). Pretonic alphabetic o in Present-day English. *Language Sciences*, (74), 1–23.
- TREIMAN, R. DECKER, K. & KESSLER, B. (2019). Adults’ sensitivity to graphotactic differences within the English vocabulary. *Applied Psycholinguistics*, (40), 167–182.
- TREIMAN, R. & KESSLER, B. (2023). Spelling-to-Sound Translation for English Disyllables: Use of Long and Short Vowels Before Single Medial Consonants. *Journal of Experimental Psychology: Learning Memory and Cognition*.
- TREIMAN, R. ROSALES, N. CUSNER, L. & KESSLER, B. (2020). Cues to stress in English spelling. *Journal of Memory and Language*, 112.
- TREVIAN, I. (1993). *Phonographématique, phonologie et morphophonologie des consonnes en anglais contemporain*. Ph.D. dissertation, Université Paris VII.

- TREVIAN, I. (2007). Stress-neutral endings in contemporary British English: an updated overview. *Language Sciences*, 29(2–3), 426–450.
- TREVIAN, I. (2015). *English Suffixes: Stress-assignment Properties, Productivity, Selection and Combinatorial Processes*. Bern: Peter Lang.
- TURCSAN, G. & HERMENT, S. (2015). Making Sense of Nonce Word Stress in English. In J. A. Mompean & J. Fouz (Eds.), *Investigating English Pronunciation: Current Trends and Directions* (pp. 23–46). Hampshire, England: Palgrave Macmillan.
- UPTON, C. & KRETZSCHMAR, W. A. (2017). *The Routledge Dictionary of Pronunciation for Current English*. The Routledge Dictionary of Pronunciation for Current English. Abingdon & New York: Routledge.
- VAN DER HULST, H. (2012). Deconstructing Stress. *Lingua*, 122(13), 1494–1521.
- VAN DER HULST, H. (2014a). Representing Rhythm. In H. Van der Hulst (Ed.), *Word Stress: Theoretical and typological issues*. (pp. 325–365). Cambridge: Cambridge University Press.
- VAN DER HULST, H. (2014b). Word Stress: Past, Present and Future. In H. van der Hulst (Ed.), *Word Stress. Theoretical and Typological Issues2* (pp. 3–55). Cambridge: Cambridge University Press.
- VAN HEUVEN, W. V. J. MANDERA, P. KEULEERS, E. & BRYLSBAERT, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, (67), 1176–1190.
- VAN SANTEN, J. P. H. (1992). Contextual effects on vowel duration. *Speech Communication*, 11(6), 513–546.
- VIDEAU, N. (2013). *Préfixation et phonologie de l'anglais: Analyse lexicographique, phonétique et acoustique*. Ph.D. dissertation, Université de Poitiers.
- VIDEAU, N. & HANOTE, S. (2015). Pronunciation of Prefixed Words in Speech: The Importance of Semantic and Intersubjective Parameters. *Lexis* [Online], 9 | 2015.
- WANG, H. S. & DERWING, B. L. (1986). More on English Vowel Shift: The Back Vowel Question. *Phonology Yearbook*, 3, 99–116.
- WATTS, V. (2004). *The Cambridge Dictionary of English Place-Names: Based on the Collections of the English Place-Name Society*. Cambridge: Cambridge University Press.
- WEIDE, R. L. (1994). CMU pronouncing dictionary. <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>
- WELLS, J. C. (1982). *Accents of English. An Introduction*. Cambridge: Cambridge University Press.
- WELLS, J. C. (1990). Syllabification and allophony. In *Studies in the pronunciation of English. A commemorative volume in honour of A. C. Gimson* (pp. 76–86). London: Routledge.
- WELLS, J. C. (2008). *Longman Pronunciation Dictionary* (3rd ed). London: Longman.
- WENNERSTROM, A. (1993). Focus on the Prefix: Evidence for Word-Internal Prosodic Words. *Phonology*, 10(2), 309–324.
- WENSZKY, N. (2004). *Secondary Stress in English Words*. Budapest: Akadémiai Kiadó.
- WIDDOWS, D. & COHEN, T. (2010). The semantic vectors package: New algorithms and public tools for distributional semantics. In *Proceedings - 2010 IEEE 4th International Conference on Semantic Computing, ICSC 2010* (pp. 9–15).
- WILSON, C. & GEORGE, B. (2009). The MaxEnt Grammar Tool [Computer Software]. <https://linguistics.ucla.edu/people/hayes/MaxentGrammarTool/>
- WURM, L. H. (1997). Auditory Processing of Prefixed English Words Is Both Continuous and Decompositional. *Journal of Memory and Language*, 37(3), 438–461.
- WURM, L. H. (2000). Auditory Processing of Polymorphemic Pseudowords. *Journal of Memory and Language*, 42, 255–271.
- YU, H. J. (2022). Geminate and singleton contrast in English affixed words. *Phonetics and*

Speech Sciences, 14(3).

ZHANG, Y. (2021). Partial dependency of vowel reduction on stress shift: Evidence from English-ion nominalization. *Proceedings of the Annual Meetings on Phonology*, 9.

ZUUR, A. F. IENO, E. N. & ELPHICK, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1).

DOCUMENT DE SYNTHÈSE

1. INTRODUCTION

Dès ma première année comme jeune chercheur au sein du groupe morphophonologie du Laboratoire Ligérien de Linguistique (ci-après LLL), j'ai appris deux choses qui sont restées centrales dans mon approche de la recherche : l'importance des **fondations empiriques** de tout modèle théorique, et l'importance d'être **ouvert à différents cadres théoriques**. La première me vient de l'approche introduite par Guierre (1979) et de comment elle s'est construite en réponse aux approches générativistes qui, pendant la deuxième moitié du XX^{ème} siècle, ont essentiellement eu recours à l'introspection ou à des exemples réutilisés d'une publication à l'autre. Comme il est courant dans l'approche guierrienne, mes travaux ont eu recours à des données dictionnaires. Ce type de données présente – à mes yeux, comme tout type de données – des avantages et des inconvénients, et j'ai donc eu recours à d'autres types de données sur certains phénomènes étudiés afin de compléter les données dictionnaires. Sur le plan méthodologique, j'ai d'abord simplement utilisé des statistiques descriptives puis me suis progressivement tourné vers des statistiques inférentielles, qui me semblent essentielles dans l'analyse de phénomènes dépendant de multiples facteurs. Le second élément essentiel de mes travaux me vient de la nécessité que je perçois de communiquer avec toutes les théories afin de faire progresser notre connaissance des phénomènes étudiés. Je crois que l'on gagne à adopter des perspectives différents et à explorer les prédictions faites par différents modèles. Cette approche est au cœur du projet ERSaF (English Root Stress across Frameworks), que je porte avec Sabine Arndt-Lappe, et dans le cadre duquel nous co-encadrons deux doctorants, Aaron Seiler et Marie Gabillet.

Si je devais définir de manière générale l'objet de mes recherches, ce serait la **phonologie des mots** (ou la phonologie lexicale), principalement en anglais britannique. Ceci concerne la distribution des accents lexicaux et la distribution des segments. Je m'intéresse aux mécanismes qui déterminent ces distributions, qu'il s'agisse de déterminants phonologiques ou de déterminants impliquant d'autres 'niveaux' ou 'modules' comme la morphosyntaxe ou l'orthographe. Cet objet de recherche m'intéresse en tant que tel, mais également en tant que 'fenêtre' sur d'autres phénomènes comme l'organisation du lexique. Je me suis donc intéressé à une variété de phénomènes liés à la phonologie des mots, et je les vois comme autant de pièces d'un puzzle extrêmement complexe. J'aborde chacun de ces phénomènes comme faisant partie d'un **système global** que je cherche à comprendre. Dans ce document, j'ai cherché à faire apparaître les liens entre ces différents phénomènes en renvoyant régulièrement d'une section vers une autre.

Enfin, je souhaiterais présenter l'esprit qui a été le mien dans la préparation de ce document et les motivations des choix qui ont été faits concernant l'organisation des différents sujets traités. J'ai choisi de rédiger ce document comme une revue critique générale des études que j'ai menées et de proposer des extensions de certaines d'entre elles. J'ai également intégré des éléments qui ont émergé dans l'année de préparation de ce document et qui n'ont pas encore fait l'objet de publications. Ces derniers seront d'une nature plus exploratoire que le reste du

document. Dans le document complet en anglais, j'ai eu à cœur de rendre mon propos accessible à des personnes qui ne sont pas nécessairement au fait de tous les concepts et théories mobilisés en intégrant des sections qui les présentent de manière synthétique. Pour ce qui est de l'organisation thématique de ces différentes études, j'ai fait le choix de traiter de chaque étude dans la section la plus pertinente, ce qui a parfois impliqué de scinder des études entre plusieurs sections. De manière générale, je vois ce document comme une vue d'ensemble des phénomènes de la phonologie des mots de l'anglais auxquels je me suis intéressé personnellement. Ceci dit, je crois que toute personne intéressée par la phonologie de l'anglais trouvera ici plusieurs questions qui y sont centrales et des résumés que j'espère utiles à leur sujet : la nature de l'organisation rythmique de l'anglais et la relation entre le segmental et le suprasegmental (§2), la nature de la morphologie et son interaction avec la phonologie (§3), le rôle de l'orthographe dans la phonologie (§4) et les divisions au sein du lexique (§5).

Le document complet fait état de nombreuses études empiriques ayant recours à des modèles statistiques. Afin de faciliter la lecture, le détail de ces tests n'est pas présenté ici et les personnes qui s'y intéresseraient sont invitées à consulter les études. Dans ce document de synthèse en français, les études sont présentées de manière plus synthétique encore, et seuls les résultats principaux sont mis en avant.

2. SYLLABES FORTES ET SYLLABES FAIBLES

Une part conséquente de mes travaux de recherche a porté sur la distribution des syllabes fortes et faibles. Il convient donc de commencer par définir les notions de « stress » et d'« accent » (j'utiliserai ici les termes « accent » et « accent mélodique » comme équivalents français des termes anglais « stress » et « accent », respectivement) et la relation qu'elles entretiennent avec la nature des voyelles. J'évoquerai également ici certains de mes travaux ayant porté sur la « réduction vocalique » (§2.3) et l'accent secondaire prétonique (§2.3) lorsque ces derniers n'interagissent pas avec la morphologie.

2.1. Accentuation et voyelles

2.1.1. Le système vocalique du SSBE

Face à l'écart grandissant entre les conventions de transcription phonémique les plus répandues, qui décrivent la Received Pronunciation (RP), et la manière dont l'anglais britannique standard du sud (Standard Southern British English, ou SSBE) est parlé de nos jours, j'explore les intérêts d'une autre analyse et d'autres conventions de transcription du système vocalique proposées par Lindsey (2019) et Szigetvári (2016, 2018). Les différences entre les deux systèmes sont présentées en (1).

(1) Les inventaires vocaliques du RP et du SSBE

	Voyelles brèves		Diphthongues		Voyelles longues			
	RP	SSBE	RP	SSBE	RP	SSBE		
KIT	ɪ	ɪ	FLEECE	i:	ɪj	NEAR	ɪə	ɪ:
DRESS	e	ɛ	FACE	eɪ	ɛj	SQUARE	eə	ɛ:
TRAP	æ	ɑ	PRICE	aɪ	ɑj	START	ɑ:	ɑ:
STRUT	ʌ	ə	MOUTH	aʊ	aw	NURSE	ɜ:	ə:
LOT	ɒ	ɔ	CHOICE	ɔɪ	ɔj	FORCE	ɔ:	o:
FOOT	ʊ	ə	GOAT	əʊ	əw	CURE ¹⁵⁸	ʊə	ə:
			GOOSE	u:	ʊw	LETTER	ə	ə

En plus de marquer des évolutions ayant affecté l'anglais britannique (notamment les changements en chaîne dans le sens « inverse des aiguilles d'une montre » dans le trapèze vocalique, Lindsey 2019 : 18), ce système représente les diphthongues comme des séquences voyelle-glide. Les modifications de notation placent clairement les voyelles de FLEECE et GOOSE dans le groupe des diphthongues. Les trois séries que forment l'ensemble des voyelles en (1) se caractérisent par des propriétés détaillées par Lindsey et rassemblées dans le Tableau 1.

		Voyelles brèves	Diphthongues	Voyelles longues
Peuvent être suivies de	V	×	✓	×
	#	×	✓	✓
	C	✓	✓	✓
	r de liaison	×	×	✓
Particulièrement sensibles au pre-fortis clipping		×	✓	×
Peuvent déclencher le pre-liquid breaking		×	✓	×
Peuvent subir le smoothing devant [ə]		×	✓	×

Tableau 1. Les propriétés des trois ensembles de voyelles selon Lindsey (2019)

Dans le document principal, je détaille ces propriétés, notamment les trois dernières, ainsi que les éléments empiriques qui permettent d'évaluer leur validité. J'en conclus que, si les données sur ces phénomènes permettant de distinguer les diphthongues des monophthongues sont peu nombreuses et partielles, celles qui sont disponibles vont globalement dans le sens des propositions de Lindsey.

Par ailleurs, la proposition d'analyser les diphthongues comme des séquences voyelle-glide permet de rendre compte de plusieurs phénomènes, notamment de la distribution de ces voyelles (elles sont les seules à pouvoir précéder d'autres voyelles) et le fait qu'elles déclenchent le « pre-liquid breaking », c'est-à-dire l'insertion d'un [ə] entre une diphthongue et une liquide (e.g. *file* ['fajəɫ]; *real* ['ri:əɫ]). Adopter ce système permet également de ne pas avoir à postuler un phénomène d'insertion de glides après les diphthongues dont le point d'arrivée est

¹⁵⁸ CURE est affiché en gris car il fusionne souvent avec FORCE (Cruttenden 2014 : 156 ; Lindsey 2019 : 48).

haut (e.g. *I am* /aɪ æm/ → [aɪ j æm]). Ces glides seraient bien présentes, mais seraient seulement une partie de la diphtongue et subiraient éventuellement une resyllabation en attaque du mot suivant (e.g. /ɑj am/ → [ɑ jam]). Selon Lindsey¹⁵⁹, cette analyse permettrait de rendre compte des différences de durées entre les diphtongues suivies d'une voyelle et les voyelles longues suivies d'une voyelle, les premières générant simplement une resyllabation de la glide (e.g. *two o'clock* /'tʌw ə'klɒk/ → ['tʰu wə'kɒk]) et les secondes entraînant l'insertion d'un [ɹ] (e.g. *four o'clock* /'fo: ə'klɒk/ → ['fo: ɹə'kɒk]).

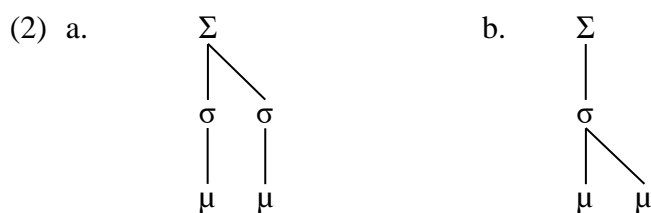
Par conséquent, dans ce document j'adopterai le système de Lindsey pour représenter les voyelles du SSBE. Par ailleurs, j'utiliserai /ɹ/ plutôt que /r/. Les transcriptions qui sont utilisées dans ce document viennent de Wells (2008) ou Jones (2006). Celles qui sont tirées de mes travaux antérieurs seront adaptées à ces nouvelles conventions de transcription, et celles tirées d'autres travaux seront adaptées également, sauf mention contraire.

2.1.2. « Stress » et « accent »

Dans la littérature, on trouve une grande confusion terminologique autour des termes renvoyant à la proéminence de certaines syllabes par rapport à d'autres : accent et accent mélodique. Par ailleurs, on trouve des différences entre les traditions britannique et américaine sur comment doivent être analysées les syllabes ayant un niveau de proéminence intermédiaire (e.g. *October*, *retrograde*). Dans cette partie, je défends l'idée que l'accent et l'accent mélodique sont des phénomènes liés mais distincts, et que les différences entre les traditions britannique et américaine ont essentiellement à voir avec le fait que la première tend à décrire les accents mélodiques alors que la deuxième décrit l'accent. On désignera sous le terme d'« accent » un type de proéminence partiellement lié à la qualité des voyelles et associé à une série de corrélats segmentaux qui seront détaillés ci-dessous. Le terme « accent mélodique » renverra quant à lui à une proéminence additionnelle pouvant être assignée à certaines syllabes accentuées. Nous verrons ci-dessous quelles syllabes sont concernées.

Dans la littérature générativiste sur la proéminence, il est courant d'utiliser le concept de pied pour rendre compte des phénomènes attribués à l'accentuation. Dans la Phonologie Prosodique, le pied est un des constituants de la Hiérarchie Prosodique avec la syllabe ou le mot phonologique, ces constituants étant les domaines dans lesquels s'appliquent les processus phonologiques et sont également des domaines de restrictions phonotactiques. Dans cette théorie, une syllabe accentuée est la tête d'un pied, et de nombreux phénomènes qui font référence aux syllabes accentuées ont été analysés en termes de structure en pieds. Il est donc utile de présenter ce concept ici car j'y aurai recours à de nombreuses reprises dans ce document. Il est courant de supposer que les pieds en anglais sont trochaïques (i.e. leur tête est à gauche) et bimoraïques (i.e. ils contiennent une syllabe lourde ou deux syllabes légères ; Bermúdez-Otero 2018b ; Bermúdez-Otero & McMahan 2006 ; Pater 2000). On aurait donc les deux structures possibles en (4), où Σ représente le pied, σ la syllabe et μ la more.

¹⁵⁹ Lindsey, Geoff, "LENGTH and LINKING in British, American and Australian accents!", mis en ligne par Dr Geoff Lindsey, 15 October 2023, <https://www.youtube.com/watch?v=tPi2jtU7Tl4>



Dans l'approche guierrienne, les pieds ne sont généralement pas utilisés, ce qui a été l'objet de critiques de la part de phonologues habitués à leur utilisation. Dans les modèles qui les utilisent, les pieds ont été utilisés pour unifier des phénomènes qui sont au cœur des études guierriennes (accentuation et distribution des voyelles), mais également pour de nombreux autres phénomènes. Ainsi, il me semble utile d'explorer les apports que peut faire cet outil théorique afin de rendre compte d'un plus large éventail de phénomènes.

Etant donné qu'indépendamment des systèmes de transcription utilisés, on trouve des voyelles qui peuvent être accentuées ou inaccentuées (e.g. c'est le cas dans tous les systèmes pour la voyelle de KIT), il est nécessaire de détailler les diagnostics pouvant être utilisés pour distinguer une voyelle accentuée d'une voyelle inaccentuée. Ces phénomènes sont identifiés dans les syllabes portant la proéminence maximale (accent primaire), et on en déduit que les autres syllabes qui présentent ces phénomènes sont également accentuées. La liste des phénomènes qui ont été présentés dans la littérature peut être trouvée dans le Tableau 2, où leur description est donnée en termes d'accent ou de structure en pieds.

Processus	Généralisation basée sur l'accent	Généralisation basée sur le pied
Aspiration de /p, t, k/ Présence de /h/ Présence de /ɪ/ dans les variétés non-rhotiques larges	En attaque de syllabe accentuée, à l'initiale d'un mot ou en attaque d'une syllabe prétonique précédée d'un pied minimal	A l'initiale d'un pied
Epenthèse entre une nasale et une fricative sourde Assimilation de /n/ devant une vélaire	Bloquée si la fricative est en attaque d'une syllabe accentuée Optionnelle si la vélaire est en attaque d'une syllabe accentuée	A l'intérieur d'un pied
Yod-dropping Allongement compensatoire	Dans une syllabe accentuée	A l'intérieur d'une projection minimale de pied
Perte de /w/ dans GOOSE	Variable si non suivi d'une syllabe inaccentuée	Variable dans la dernière syllabe dépendante d'un pied
Syncope	Bloquée devant une syllabe accentuée	Si la syncope améliore la structure en pieds
Smoothing	Si la première voyelle est accentuée	Si la première voyelle appartient à la tête du pied
Formation de consonne syllabique	Bloquée dans les syllabes accentuées (sauf pour /ɪ/ en GA)	Interdite pour la tête d'un pied

Tableau 2. Les diagnostics segmentaux de l'accent proposés dans la littérature pour l'anglais

Là encore, il est important de souligner que beaucoup de ces propositions s'appuient sur une base empirique assez restreinte, souvent appuyée sur l'introspection, et que des études pour approfondir notre connaissance de ces phénomènes seraient les bienvenues. Ceci étant dit, les éléments disponibles montrent qu'il est bien possible d'avoir des accents subsidiaires, qu'on appellera « secondaires », post-toniques, i.e. à la suite de l'accent principal.

Plusieurs études phonétiques récentes ont montré que les dérivés ou les composés comportant un accent primaire et un accent secondaire se comportent différemment selon l'ordre dans lesquels ils se trouvent. Si l'accent primaire est premier (e.g. *windshield*, *randomize*, *activate*) alors on n'a qu'un seul « accent » sur cette syllabe et aucun sur l'accent secondaire post-tonique, alors que si les accents se trouvent dans la configuration inverse (e.g. *Park Avenue*, *violation*, *published*), on a un accent mélodique sur chaque syllabe accentuée. Dans le document principal, plusieurs autres configurations sont évoquées, et il semble que la situation soit légèrement plus complexe que cela puisque, comme le montre Fox (2000), le nombre d'accents mélodiques qui sont réalisés dépend du contexte d'énonciation. Il défend alors la nécessité de distinguer les accents mélodiques réalisés des accents mélodiques potentiels. Les résultats phonétiques évoqués ci-dessus suggèrent que, dans certaines configurations, des schémas dominants se dégagent, mais ceux-ci ne sont pas forcément systématisables à tous les cas de figure.

Il faut évoquer ici le fait que les dictionnaires de prononciation comme Wells (2008) et Jones (2006) – références dans les concours de l'enseignement secondaire français – sont peu cohérents dans l'utilisation des « stress marks ». En effet, ces marques sont utilisées dans certains cas pour les syllabes post-toniques (suffixe *-ism*, composé dont le deuxième constituant est dissyllabique), de telle sorte qu'on ne peut pas dire s'ils indiquent l'accent ou l'accent mélodique. De manière générale, ils sembleraient favoriser la notation des accents mélodiques habituellement réalisés (comme le défendent Abercrombie (1976) et Schane (2007)) mais ces cas d'accents post-toniques y font exception puisque les syllabes concernées ne reçoivent pas habituellement d'accents mélodiques.

Enfin, je termine cette section en évoquant le lien entre structure segmentale et accent. Dans la tradition générativiste, il est généralement supposé que la structure de la rime détermine la position des accents. Cet élément était central dans le modèle de Chomsky & Halle (1968, ci-après *SPE*) – qui parlaient alors de noyaux lourds ou légers plutôt que de structure syllabique – et a fait l'objet, d'une part, de fortes critiques dans le courant guierrien et, d'autre part, de nuances dans les travaux générativistes à la suite de *SPE*. L'influence des syllabes lourdes a été supposée variable dans certaines configurations (notamment si la coda est /s/ ou une sonante), et avec l'avènement de la Théorie de l'Optimalité (Optimality Theory, ci-après OT ; Prince & Smolensky 1993) elle est devenue une contrainte qu'il est possible d'enfreindre (voir Pater 2000). Par ailleurs, il a été proposé que l'autre composante de la syllabe, l'attaque, puisse affecter le placement de l'accent, au moins dans les dissyllabes (Kelly 2004 ; Ryan 2014).

Ainsi, dans les modèles les plus répandus, on suppose que la position des accents dépend du poids des syllabes, ce qui inclut les voyelles. Par exemple, dans *horizon* /hə'ɹɔjzən/, on suppose que le « point de départ » est [həɹɔjzən], et que c'est la présence de [ɹj] dans la pénultième qui permet de prédire la position de l'accent (ici pénultième). Or, l'apprenant rencontre généralement les mots nouveaux avec les deux informations (voyelles et accent) disponibles (à l'oral) ou sans aucune des deux (à l'écrit). La tradition guierrienne a exploré en

détail les règles de lecture qui permettent d'établir la position de l'accent et la réalisation des voyelles à partir de la représentation orthographique, et permet donc de dégager les éléments de la structure segmentale qui sont utiles pour assigner l'accent à un mot inconnu sur la base d'un système phonologique ayant accès à l'information orthographique : certaines séquences de consonnes dans des positions spécifiques, les digraphes vocaliques et la présence de diphtongues dans certains dérivés suffixaux.

2.2. La réduction vocalique

La « réduction vocalique » renvoie à des affaiblissements vocaliques dans des positions faibles, que l'on peut voir comme une forme de désaccentuation ou un processus allant de pair avec la désaccentuation. Ainsi, les termes de voyelle « pleine » et « réduite » qui seront utilisés ici pourraient être substitués, respectivement, par « accentuée » et « inaccentuée ». La réduction vocalique est ici vue comme un phénomène partiellement conditionné lexicalement mais qui doit faire partie de la compétence des usagers de l'anglais.

C'est un sujet sur lequel j'ai travaillé, dans un premier temps, sur ce qui est parfois appelé l'« Arab Rule », qui rend compte de deux prononciations idiolectales trouvées en anglais américain pour le mot *Arab* : /'æɪəb/ and /'ɛɪɹəb/. Cette règle dicte que les voyelles trouvées dans des syllabes fermées par une consonne non-coronale tendent à ne pas être réduites, sauf si la syllabe concernée est précédée par une syllabe légère (Fidelholtz 1966 ; Hayes 1982 ; Pater 1995, 2000 ; Ross 1972). J'ai mené une étude dictionnaire qui a permis de montrer que le phénomène était bien attesté en anglais britannique également, même lorsque la fréquence lexicale des mots est prise en considération. Ces résultats ont été publiés dans un article coécrit avec Guillaume Enguehard, Nicola Lampitelli et Jean-Michel Fournier (Dabouis *et al.* 2020), dans lequel nous développons également une analyse du phénomène dans le cadre de la Phonologie du Gouvernement.

L'analyse proposée faisait des prédictions pour d'autres configurations que celle de l'Arab Rule. Ainsi, sachant que j'avais commencé une revue de littérature importante sur le sujet, j'ai continué à travailler sur le sujet avec Jean-Michel Fournier. La littérature fait état des propositions rassemblées dans le Tableau 3 concernant les déterminants possibles de la réduction.

		Favorise la réduction	Défavorise la réduction
Structure syllabique		Syllabe ouverte	Syllabe fermée
Nature de la coda	Lieu d'articulation	Coronale	Non-coronale
	Manière d'articulation	Sonante	Obstruante
Arab Rule		Syllabe précédente légère	Syllabe précédente lourde
Position		Non-initiale	Initiale
Traits vocaliques	Position prévocalique	Relâchée	Tendue
	En fin de mot	Basse	Non-basse
Fréquence		Fréquence élevée	Fréquence basse
Caractère étranger		Natif	Etranger
Orthographe		Monographe	Digraphe
Morphologie	Dérivé à déplacement accentuel	Non-dérivé / pas de déplacement accentuel	Syllabe accentuée dans la base
	Statut de la syllabe initiale	Préfixe opaque	Partie de la racine
	Statut de la syllabe finale	Syllabe non-unique de la racine	Syllabe unique de la racine

Tableau 3. Les déterminants de la réduction vocalique

La littérature ne comportant pas d'étude large permettant d'établir le rôle de ces différents déterminants et leurs éventuelles interactions, nous avons ainsi mené une vaste étude dictionnaire, qui sera publiée prochainement dans *Phonology*. Nous nous sommes concentrés sur deux positions, la position initiale prétonique (e.g. *arrive*, *dextérité*, *hérétique*) et la position intertonique (e.g. *notoriété pionnier*, *relaxation*) ainsi que sur trois groupes de mots : des mots que nous avons qualifiés de « non-dérivés » (i.e. sans structure interne ou constitués d'une racine liée et d'un suffixe, e.g. *acacia*, *tarantula*, *ambition*), des préfixés à préfixe monosyllabique pouvant être plus ou moins transparents (e.g. *believe*, *collect*, *reactivate*, *unaltered*) et des suffixés à déplacement accentuel dans lesquels la base est accentuée sur la deuxième syllabe (e.g. *vital* ♦ *vitálicity*; *inform* ♦ *informátion*). Le corpus complet comporte 4494 mots tirés de Jones (2006) et Wells (2008) à l'aide de la Base de Données Dictionnaires du LLL et annotés manuellement. Les données ont été analysées à l'aide de modèles de régressions statistiques ordinales dans R (R Core Team 2023, v.4.1.3). Les résultats principaux pour ce qui concerne les mots non-dérivés sont les suivants : les voyelles pleines sont plus courantes si

- la voyelle est un digraphe (e.g. *audition* /o:'dɪʃən/, *araucaria* /,a.ɪo:'kɛ:ɪjə/) que s'il s'agit d'un monographe (e.g. *falafel* /fə'la:fəl/, *pioneer* /,pajə'ni:ə/);
- la voyelle est dans une syllabe fermée (e.g. *campaign* /kəm'peɪn/, *chimpanzee* /,tʃɪmpən'zi:/) que si elle se trouve dans une syllabe ouverte (e.g. *gorilla* /gə'ɹɪlə/, *propaganda* /,prɒpə'gandə/). Les voyelles suivies par /sC/ ont un comportement intermédiaire (e.g. *cascade* /kas'keɪd/, *balustrade* /,balə'streɪd/);

- le mot a été catégorisé comme étranger (e.g. *ménage*, *chandelier*) que s'il ne l'a pas été (e.g. *molest*, *incantation*);
- le mot a une fréquence plus basse que si sa fréquence est plus élevée;
- dans la position intertonique, si la syllabe précédente est lourde (e.g. *ecraseur* /ɛjkɾɑ:'zə:/, *trampoline* /tɾampə'lijn/) que si elle est légère (e.g. *magazine* /magə'zi:n/, *memorandum* /memə'ɾandəm/).

Dans des modèles complémentaires, nous avons trouvé que les différents monogrames avait un comportement différent, <a> et <e> représentant plus souvent des voyelles réduites que <o> et <u>.

Nos résultats confirment donc un certain nombre de propositions trouvées dans la littérature et rassemblées dans le Tableau 3. Le seul effet non confirmé dans les non-dérivés est l'effet de la nature de la coda. Les résultats concernant les autres catégories de mots étudiées seront mentionnés en §3.2 pour les préfixés et en §3.5 pour les suffixés.

2.3. L'accent secondaire prétonique dans les mots sans structure interne

L'accent secondaire prétonique a été l'objet de mon travail de thèse et de plusieurs publications à sa suite. Un des objectifs généraux de ce travail était de tester empiriquement les propositions trouvées dans la littérature au sujet du placement des accents secondaires prétoniques. Cette section se focalise sur les mots sans structure interne, pour lesquels le placement des accents secondaires dépend essentiellement de contraintes rythmiques et de la structure syllabique des premières syllabes. Au moment de mener cette étude, j'avais alors adopté une définition de l'accent différente de celle vue au §2.1.2, qui avait pour conséquence de ne pas traiter comme accentuées certaines syllabes à voyelles pleines.

Si l'on réanalyse les données avec cette nouvelle approche, on observe tout d'abord que les mots pour lesquels l'accent primaire est sur la troisième syllabe ont très majoritairement le schéma /201(-)/, et que les cas de schéma /221(-)/ sont plus courants dans le vocabulaire étranger. Pour les mots dont la séquence prétonique fait plus de deux syllabes, les tendances générales restent les mêmes également : l'accent secondaire est généralement placé sur la première syllabe, mais on trouve une part importante de variation, les autres schémas étant essentiellement attribuables à la position des syllabes lourdes dans la séquence prétonique et à la présence ou l'absence d'attaque initiale.

3. L'INTERACTION ENTRE LA MORPHOSYNTAXE ET LA PHONOLOGIE

3.1. Approche de la morphologie

Le cadre dans lequel j'ai été formé est, comme de nombreux modèles en phonologie, un modèle supposant l'existence d'entités morphologiques sous le niveau du mot. J'ai été amené à me renseigner sur les différents modèles théoriques en morphologie et donc sur les problèmes posés par la conception traditionnelle du morphème comme unité minimale significative. Se pose donc la question des unités de base de l'analyse morphologique.

Toutes les théories reconnaissent la difficulté à définir de manière universelle la notion de « mot ». Elles utilisent cependant deux concepts qui renvoient à ce que l'on désigne généralement comme « mot » (et j'utiliserai d'ailleurs « mot » dans les cas où la distinction entre les deux n'est pas indispensable) :

- le **lexème**, qui renvoie à une unité abstraite rassemblant plusieurs mots-formes partageant le même sens lexical ;
- le **mot-forme**, qui renvoie à une forme phonologique ou orthographique, indépendamment de son sens.

Ainsi, le lexème BE a plusieurs mots-formes : *am, is, are, was, were, being* ou encore *been*.

Les théories divergent également au sujet des unités en dessous du niveau du mot. Dans les manuels, on donne généralement le **morphème** comme unité de base de la morphologie, défini comme la plus petite unité significative. Il s'agirait d'une entité abstraite dont les réalisations concrètes s'appellent des **morphes**. Dans les cas où un morphème se réalise par plusieurs morphes différents, on parle alors d'**allomorphes**. Au sein de ces allomorphes, on peut distinguer ceux qui sont prévisibles sur des bases phonologiques (e.g. /{z, s, ɪz}/ pour le pluriel -s) de ceux qui ne le sont pas (e.g. *good – better – worse*). Dans ces cas, on parle de **supplétion**, mais la frontière entre les deux est fortement dépendante de la théorie phonologique adoptée.

Cependant, le morphème a été fortement critiqué dans la littérature, de sorte que la plupart des morphologues pratiquent aujourd'hui une morphologie dite « word-based », dans laquelle le morphème n'a au mieux qu'un statut secondaire. Par ailleurs, certaines théories récentes supposent que les unités de la morphologie sont gradientes et que le morphème 'prototypique' n'est qu'une des associations possibles entre la forme et le sens. Pour désigner les unités sous le niveau du mot, on utilisera alors le terme de **formant**, qui inclut les morphèmes et des éléments participant à la construction des mots mais dont le sens est obscur (Bauer *et al.* 2013 : 16).

Parmi ces formants, on trouve les **racines**, généralement définies comme des unités minimales sans catégorie syntaxique pouvant être libres ou liées. On y inclura des formants dont le sens est difficilement accessible (e.g. *duce* dans *induce* ou *reduce*). On trouve également le terme de **thème** (« stem » en anglais), dont les définitions sont très variables dans la littérature. La définition la plus courante est qu'il s'agit de la partie d'un mot-forme restante si

l'on retire les affixes flexionnels. Je n'utiliserai ce terme qu'en renvoyant à des travaux qui l'utilisent, ce qui pourra renvoyer à trois usages :

- Dans la littérature en psycholinguistique, il renvoie généralement à des racines ;
- En Phonologie Stratale, il est défini comme une unité lexicale ayant une catégorie syntaxique définie (Bermúdez-Otero 2018b) ;
- Raffelsiefen l'utilise pour renvoyer à la partie restante d'un mot complexe morphologiquement si l'on retire un ou plusieurs affixes (qui peuvent être des affixes dérivationnels, Raffelsiefen 2023).

Enfin, les **affixes** sont des formants liés qui s'attachent aux racines. Certains sont analysables comme des morphèmes et d'autres ont un sémantisme plus obscur (e.g. *con* dans *contain* ou *construct*).

Il nous faut aussi présenter les concepts relatifs aux relations morphologiques entre mots. Je me concentrerai ici sur les mots affixés, qu'on verra comme résultant de la **dérivation**. Ce à quoi est attaché un affixe sera appelé la **base** et le résultat de la dérivation le **dérivé**. Pour renvoyer aux classes de mots partageant des éléments morphologiques, on n'utilisera pas le terme de paradigme, qui est trop général. En suivant la littérature en morphologie, on utilisera plutôt le terme de **catégorie morphologique** pour désigner l'ensemble des mots dérivés avec le même affixe ou par le même processus morphologique, et de **famille morphologique** d'un mot pour désigner tous les mots complexes dans lesquels ce mot apparaît comme élément constitutif.

On s'intéressera notamment aux relations de **dépendance paradigmatique**, qui désignent les cas où la forme d'une expression linguistique est prévisible à partir de la représentation de surface d'une ou plusieurs expressions liées morpho-syntaxiquement (Bermúdez-Otero 2016). Le mot servant de modèle dans ces relations est désigné comme « base phonologique », et il y a controverse sur le choix des mots pouvant être une base phonologique (§3.6). On distinguera alors la **base locale** pour désigner le mot qui est un sous-constituant immédiat du dérivé considéré des **bases non-locales** qui ne le sont pas. Parmi celles-ci, on trouve les **bases distantes**, qui sont 'contenues' dans le dérivé (e.g. *direct* ♦ *director* ♦ *directorial*) et les **co-dérivés**, qui sont des lexèmes membres de la famille morphologique du dérivé considéré mais ne sont pas contenus dans celui-ci. Ces termes sont illustrés dans la Figure 47.

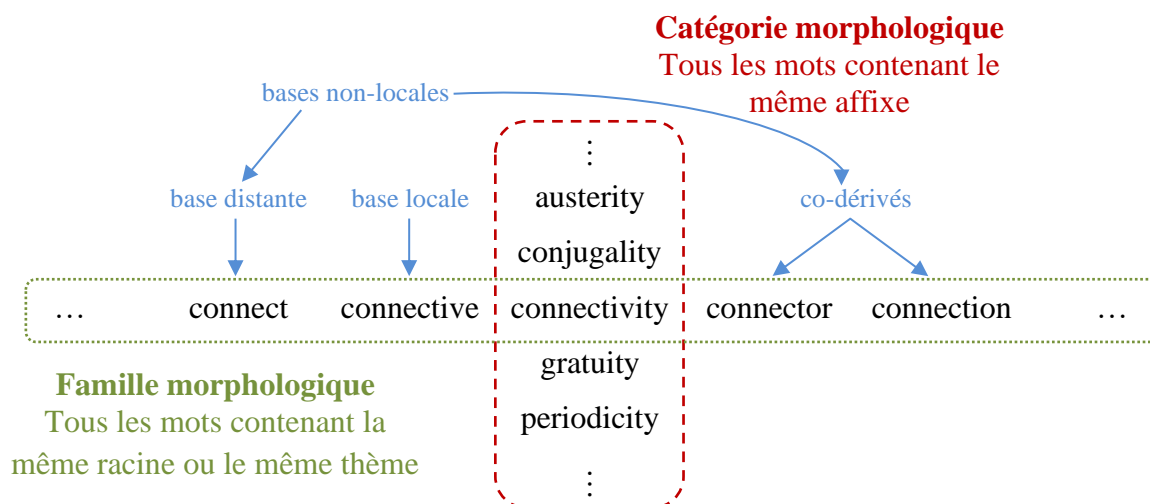


Figure 47. Illustration de la terminologie utilisée pour renvoyer aux relations morphologiques entre mots

Pour finir, à la suite de Bauer *et al.* (2013), j'utilise le symbole ♦ pour indiquer les relations de parenté morphologique sans indiquer une direction particulière.

3.2. Morphologie opaque et phonologie

Une partie de mes travaux porte sur les manifestations phonologiques de mots dont les formants sont sémantiquement obscurs. Je me suis intéressé aux préfixés opaques, aux noms propres complexes, à la question de l'apprenabilité de telles structures et à la nature des connaissances que les usagers de l'anglais peuvent acquérir à leur sujet.

Dans un article récent écrit avec Jean-Michel Fournier (Dabouis & J.-M. Fournier 2024), nous rassemblons l'ensemble des éléments nous renseignant sur le statut des préfixés historiques tels que *contain*, *declare* ou *submit*, dont les constituants peuvent être récurrents mais dont le sens n'est pas transparent. Sur le plan phonologique, on trouve que ces mots se distinguent des mots non-préfixés comme suit :

- La position de l'accent primaire diffère significativement entre les préfixés opaques et les non-préfixés, quelle que soit la catégorie, mais cette différence est plus marquée pour les non-substantifs (verbes, adjectifs et adverbes) ;
- Dans les paires verbe-nom dissyllabiques (e.g. *convict*, *concrete*, *exile*), les paires partageant un préfixe tendent à évoluer dans la même direction pour l'accentuation, ce qui suggère qu'elles partagent un formant morphologique (Sonderegger & Niyogi 2013) ;
- La réduction vocalique est quasi-systématique en position initiale prétonique si cette syllabe est un préfixe opaque (e.g. /ə/dvʌ́nce, c/ə/ndé́nse, s/ə/btráct). Cet effet prévaut sur les effets de la structure syllabique, et nous avons confirmé cet effet dans notre article sur la réduction (Dabouis & J.-M. Fournier à paraître) ;

- A l'inverse, les préfixés opaques dissyllabiques accentués sur leur première syllabe tendent à avoir une voyelle pleine dans leur deuxième syllabe contrairement aux non-préfixés (e.g. *cóntr/ɑ:/st* cp. *báll/ə/st*) ;
- Les préfixés opaques favorisent la Préservation Faible (i.e. la reprise de l'accent primaire de la base sous la forme d'un accent secondaire), comme nous l'avons montré dans Arndt-Lappe & Dabouis (soumis) ;
- On trouve des agrégats consonantiques médians dans les préfixés opaques qu'on ne retrouve pas dans les mots sans structure interne (e.g. /bs/ in *absence*).

Dans notre article, nous présentons aussi des arguments complémentaires, notamment issus de la psycholinguistique, qui montrent que les préfixés opaques sont traités différemment des mots sans morphologie par les usagers de l'anglais, que ce soit dans des tâches de décision lexicale ou des études sur la lecture de dissyllabes. Par ailleurs, nous montrons que ces constructions ne sont pas toutes entièrement opaques, puisque l'on trouve des éléments de sens partagés par des mots partageant un formant (e.g. *rupt* 'breaking', *re-* 'back' dans *rebound* ou *repel*) et que certaines racines portent même de « l'information thématique squelettique » (Baeskow 2006).

Pour les noms propres, j'ai récemment publié un article (Dabouis 2023d) dans lequel j'étudie huit caractéristiques phonologiques permettant de distinguer les mots à morphologie « analytique » (J. Kaye 1995), c'est-à-dire les suffixés neutres et les composés, des mots sans structure morphologique. Je me suis intéressé à des noms de lieu ou de famille contenant des éléments récurrents identifiés par les onomasticiens (e.g. *Washington*, *Cambridge*) et ai constitué un jeu de données de 3579 noms propres complexes dont les prononciations sont tirées de Wells (2008). Les résultats montrent que plus de la moitié de ces noms présentent des caractéristiques phonologiques qui sont typiques des mots complexes, bien que le sens de ces noms propres soit opaque (e.g. *Appleby* (nom de ville ou de famille) n'a aucun lien avec *apple*). Ceci suggère que la structure de ces mots est pertinente pour la phonologie, puisqu'ils semblent être traités comme des composés ou des mots à suffixe neutre. Certains éléments de ces résultats sont particulièrement intéressants :

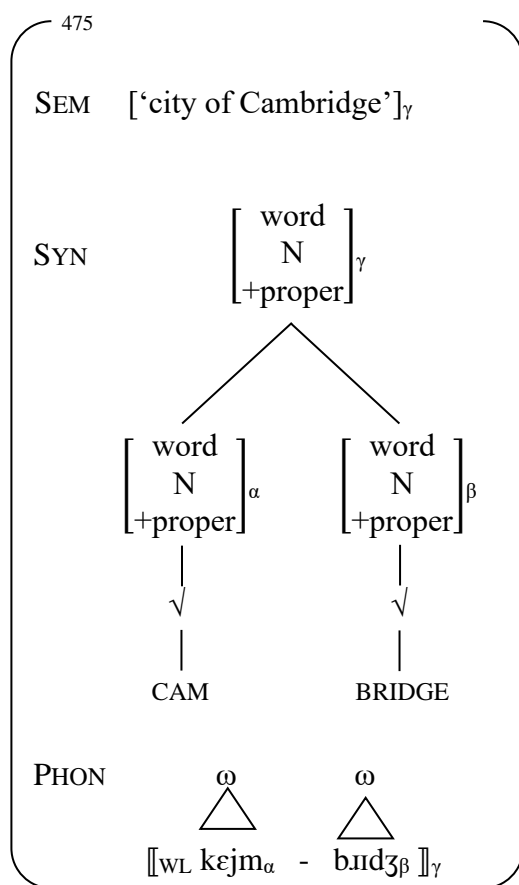
- Certaines caractéristiques des mots complexes sont plus courantes lorsque le premier formant est homographe d'un mot libre, même sans lien sémantique avec ce mot ;
- On trouve un effet de la récurrence sur la réduction du deuxième élément de ces noms : plus un formant est récurrent, plus il tend à être réduit, ce qui est cohérent avec les effets connus de la fréquence sur la réduction.

Concernant l'apprenabilité de telles structures sémantiquement appauvries, nous suggérons que cette capacité se développe tard, la littérature sur l'acquisition de la morphologie dérivationnelle montrant que les affixes acquis en premier par les apprenants de l'anglais comme langue maternelle sont ceux pour lesquels la forme et le sens sont constants, ceux associés à une forme variable (notamment de la base) étant acquis plus tard. Au-delà des phénomènes phonologiques que nous venons d'évoquer qui peuvent signaler la présence de structure morphologique, nous suggérons, à la suite de plusieurs auteurs, que le mécanisme de la commutation est au cœur de ce qui permet la détection de structures sémantiquement appauvries. C'est la récurrence des unités et la commutation croisée des préfixes et des racines

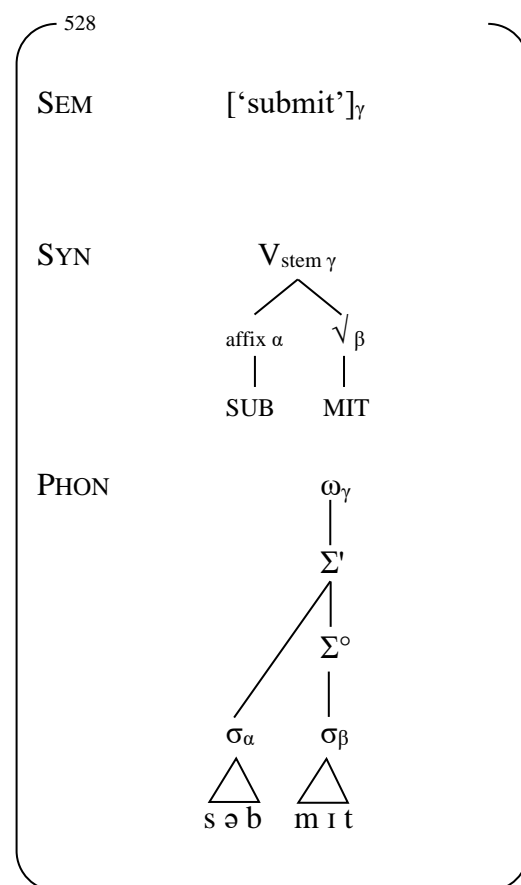
qui permet vraisemblablement la détection des formants les plus opaques tels que *se-* (e.g. *secede* cp. *accede*, *concede*, *precede*).

Enfin, nous suggérons que les usagers analysent ces mots comme étant complexes morphologiquement, ce qui peut se manifester par des représentations phonologiques typiques des mots complexes, mais ne correspond pas nécessairement à une complexité au niveau sémantique. Par exemple, les entrées lexicales de mots comme *Cambridge* ou *submit* seraient initialement de la forme de (47), où SEM représente le sens, SYN la morphosyntaxe et PHON la phonologie. A la suite de Jackendoff (1997, 2002), on suppose que les différents niveaux de représentation peuvent être coindexés, ce qui est représenté par les lettres grecques souscrites (α , β , γ).

(3) a. Lexical entry for *Cambridge*



b. Lexical entry for *submit*



Ici, crucialement, le sens est coindexé uniquement au niveau du mot entier et non au niveau de formants sous le niveau du mot. Cependant, si un usager de l'anglais apprend suffisamment de mots contenant ces formants et y détecte un sémantisme récurrent (par exemple, le sens 'back' chez certains préfixés en *re-*), alors des éléments de la représentation sémantique pourraient être coindexés avec les formants concernés.

Cette thématique nécessiterait d'être encore étudiée en diversifiant les outils utilisés, et il pourrait notamment être intéressant d'utiliser des modèles computationnels comme les modèles analogiques (Arndt-Lappe 2015 ; Skousen 2002), le Naïve Discriminative Learning (Baayen *et al.* 2011) ou encore la Sémantique Distributionnelle (Widdows & Cohen 2010), qui permettraient de nous éclairer sur l'apprenabilité de telles structures.

3.3. Accentuation et catégorie syntaxique

La littérature en phonologie anglaise fait état de différences accentuelles fondées sur la catégorie syntaxique des mots, et on se concentrera ici sur les verbes et les noms. Pour certains, la différence entre ces deux catégories est attribuée à une règle d'extramétricalité (i.e. invisibilité de matériau phonologique) différente : dernière consonne pour les verbes contre dernière syllabe pour les noms. Cette différence implique que les effets de poids syllabique se manifestent différemment dans ces deux catégories. D'autres attribuent la différence à une sensibilité différente à la préfixation opaque (voir la section précédente). Enfin, d'autres supposent que les catégories en tant que telles seraient différentes, notamment les dissyllabes, pour lesquels les verbes préféreraient l'accentuation finale et les noms l'accentuation initiale. Dans le cadre du projet ERSaF, nous avons passé en revue les études expérimentales et sur dictionnaire et avons trouvé que des effets de poids syllabique, de la préfixation et de la catégorie syntaxique sont tous confirmés lorsqu'ils sont testés, mais que ces facteurs sont rarement testés en même temps.

Avec Jean-Michel Fournier (aidés par d'autres collègues), nous avons mené une étude large de l'accentuation des verbes sur l'ensemble de ceux qui sont répertoriés dans Jones (2006), soit 3548 verbes après nettoyage (entrées obsolètes, conversions). Cette étude a été publiée dans un ouvrage collectif sur l'accentuation anglaise (Dabouis & J.-M. Fournier 2023). Nous avons montré un aspect intéressant de la structure morphologique des verbes : ceux-ci sont fréquemment construits « à gauche » (préfixation ou particule adverbiale), ce qui pourrait expliquer le fort effet de la préfixation sur leur accentuation. Notre étude détaille le comportement accentuel par catégorie morphologique, où l'on quantifie et confirme les effets des suffixes affectant l'accentuation (e.g. *-ate*, *-ify*) ou neutres (e.g. *-en*), nous proposons la seule analyse de l'accentuation des composés verbaux à notre connaissance (e.g. *frostbite*, *waterlog*, *handcraft*), et nous analysons de près l'accentuation des verbes sans structure interne (e.g. *argue*, *follow*, *gallivant*) ou préfixés opaques (e.g. *commit*, *eject*, *remember*). Pour ces deux dernières classes, nous avons montré que la préfixation opaque et le poids de la dernière syllabe prédisent efficacement la position de l'accent et se recouvrent largement. Nous suggérons que les deux sont nécessaires pour rendre compte de l'accentuation des verbes.

Dans le cadre du projet ERSaF, Aaron Seiler a mené une étude d'élicitation de pseudo-mots trisyllabiques présentés en tant que verbes. Les variables d'intérêt étaient le poids des syllabes (les pseudo-mots avaient une seule syllabe lourde dans l'une des trois positions possibles) et la préfixation (ils pouvaient commencer par une séquence graphique trouvée pour un préfixe productif, non-productif ou aucun préfixe). Les résultats montrent une prédominance de l'accentuation antépénultième dans les productions des participants bien supérieure à celle observée dans les verbes comparables dans le lexique. Les effets du poids syllabique sont bien présents mais divergent de ce qui est trouvé dans la littérature puisque le poids joue un rôle dans toutes les positions, et pas seulement dans la position finale (même si c'est là où son effet est le plus fort). Concernant la préfixation, les effets sont mitigés puisque l'on observe des différences significatives pour les pseudo-verbes commençant par une séquence correspondant à un préfixe productif (e.g. *pre-*, *re-*, *un-*), mais pas pour ceux correspondant à des préfixes non-productifs

(e.g. *be-*, *con-*, *se-*). Cependant, on observe des disparités importantes au sein de ces deux groupes, qui suggèrent qu'une analyse par préfixe serait nécessaire plutôt que des regroupements basés sur la productivité, mais des données supplémentaires seraient nécessaires car celles-ci ne comportent pas assez d'items différents par préfixe. Des premières explorations ont cependant été menées en utilisant la mesure de « prefix likelihood » proposée par Wurm (2000), qui représente la proportion de mots préfixés parmi les mots commençant par la même séquence graphique : les séquences correspondant plus souvent à des préfixes sont moins souvent accentuées que celles qui correspondent moins souvent à des préfixes.

J'ai personnellement peu travaillé sur l'accentuation des noms, mais Marie Gabillet, travaille actuellement dessus dans le cadre du projet ERSaF. Son travail devrait permettre d'éclairer le rôle du poids syllabique (attaques de syllabe incluses), celui des sous-lexiques, de la préfixation opaque, de la longueur des mots et des noyaux finaux. De premières explorations sur le rôle des sous-lexiques et des noyaux finaux seront évoquées au §5.

Toujours dans le cadre du projet ERSaF, nous avons mené une étude de jugements sur des non-mots dissyllabiques présentés auditivement et visuellement en contextes verbaux ou nominaux. Cette étude avait initialement été conçue dans le cadre d'un mémoire de master pour un étudiant que j'encadrais, Youssef Fadel. L'objectif était d'évaluer le rôle du nombre de consonnes en attaque initiale et en coda finale, de la préfixation et de la catégorie syntaxique. Les résultats montrent que les participants ont choisi une proportion importante d'accentuation finale en comparaison des taux d'accentuation finale observés dans les dictionnaires. On trouve également des effets de tous les facteurs étudiés dans les directions attendues, même s'ils sont assez limités. Là encore, les différentes séquences utilisées qui correspondent à des préfixes ne se comportent pas toutes de la même manière, et des tests préliminaires avec la « prefix likelihood » montrent des effets prometteurs de cette variable.

Enfin, la dernière thématique abordée ici relative au lien entre catégorie syntaxique et accentuation concerne les paires dissyllabiques verbe-nom, qui était un des premiers sujets sur lesquels j'ai travaillé. Il s'agit d'un des projets collectifs du groupe morphophonologie du LLL et auquel j'ai participé entre 2012 et 2014. Cette classe de mot est bien connue des spécialistes de l'anglais puisque certains de ces paires sont invoquées pour suggérer que l'accentuation pourrait être distinctive, certaines paires se distinguant uniquement par l'accentuation : *increase*_N /'m,kɪŋs/ cp. *increase*_V /m'kɪŋs/. En réalité, les études existantes montrent trois comportements accentuels possibles pour ces paires :

- (4) a. **Verbal** (oxytonique ; ≈ 60%) : le nom et le verbe sont accentués sur leur deuxième syllabe (e.g. *concern*, *eclipse*, *mistake*, *supply*);
- b. **Alternant** (diatonique ; ≈ 30%) : le verbe est accentué en finale et le nom à l'initiale (e.g. *addict*, *import*, *permit*, *record*);
- c. **Nominal** (paroxytonique ; ≈ 10%) : le nom et le verbe sont accentués sur leur première syllabe (e.g. *contact*, *distance*, *profile*, *surface*).

La question qui n'a pas trouvé de réponse dans la littérature existante est celle des déterminants qui font qu'une paire donnée appartient au type verbal, alternant ou nominal. Nous avons donc

étudié un inventaire de 186 paires verbe-nom préfixées (opaques)¹⁶⁰ avec des données tirées de trois dictionnaires de prononciation pour trois variétés d'anglais (britannique, américain, australien). Nous avons observé une importante variation accentuelle au sein de chaque variété d'anglais (une paire sur cinq), mais une forte stabilité entre les variétés. Nos études de cet inventaire ont montré des effets de la fréquence relative des membres de la paire, du poids de la deuxième syllabe et des dates de première attestation dans l'*Oxford English Dictionary* (ci-après *OED*). Cependant, ces études étaient fondées sur des analyses de chaque facteur pris de manière isolée, et les interactions entre facteurs n'ont pas été considérées. Castanier (2023) suggère qu'un élément qui pourrait être déterminant est la nature de la relation sémantique entre le nom et le verbe. Un des étudiants que j'encadre en master, Julien Penard, travaille actuellement sur le sujet.

3.4. Les classes de suffixes

Les suffixes de l'anglais ont des effets variables sur l'accentuation : certains peuvent entraîner des « déplacements accentuels » entre base et dérivé (e.g. *héro* ♦ *heróic*, *módern* ♦ *modernity*), d'autres sont accentuellement « neutres » et peuvent ainsi entraîner des schémas accentuels inhabituels comme l'accentuation pré-antépénultième (e.g. *advénturousness*, *cháacterless*, *síimilarly*), et d'autres oscillent entre les deux (e.g. *vólcanist* ♦ *volcáno*, *deféatist* ♦ *deféat*). Ce sujet est le premier sur lequel j'ai travaillé, mon mémoire ayant porté sur l'accentuation des adjectifs suffixés en *-al*. Nous avons également travaillé sur le sujet avec Marjolaine Martin et avons fait une présentation synthétisant les différents traitements qui ont été proposés sur ces classes de suffixes (Dabouis & M. Martin 2022). Une synthèse de ces différentes analyses peut être trouvée dans le Tableau 4 sur la base de deux critères :

- le comportement de ces classes de suffixes est-il attribué à des différences **procédurales** (i.e. computationnelles) ou **représentationnelles** (e.g. forme phonologique des affixes, prosodification) ?
- le comportement de ces classes de suffixes est-il supposé être **arbitraire** ?

¹⁶⁰ Nous nous sommes concentrés sur cette catégorie car c'est la principale concernée par ces différents comportements accentuels.

	Outil procédural	Outil représentationnel	Comportement supposé arbitraire
<i>SPE</i>	Le cycle	Frontières # +	Oui
Ecole de Guierre (J.-M. Fournier)	Mode de calcul / cophonologie	Aucun	Oui
Phonologie Lexicale / Stratale	Affiliation stratale des suffixes Le cycle	Aucun	Oui, mais en partie lié à la nature des bases
OO-Correspondence (Benua 1997)	Contraintes de correspondance Output-Output	Aucun	Oui
Raffelsiefen (2005, 2023)	Contraintes de correspondance Output-Output	Cohésion	Partiellement, les suffixes à consonne initiale sont neutres
Newell (2021)	Le cycle (phases)	Voyelles flottantes	Non

Tableau 4. Résumé des différentes analyses des comportements phonologiques de différents suffixes

Ce qui ressort d'une telle revue de littérature est que les classes de suffixes vont bien au-delà de l'accentuation et se manifestent sur le plan segmental et dans l'implémentation phonétique. Il semble souhaitable de développer des analyses ayant recours à la fois à des outils procéduraux et représentationnels, notamment pour rendre compte des comportements des suffixes à consonne initiale relevés par Raffelsiefen (2005, 2023). Ceci peut être fait par exemple dans un cadre stratal en supposant une prosodification différente pour les suffixes selon qu'ils commencent par une voyelle ou une consonne.

3.5. L'hypothèse de la segmentabilité

Plusieurs études auxquelles j'ai contribué ont cherché à évaluer (entre autres choses) « l'hypothèse de la segmentabilité », initialement proposée par Hay (2001, 2003). Cette hypothèse s'appuie sur des modèles psycholinguistiques de l'accès lexical qui supposent que les mots complexes peuvent être récupérés dans le lexique mental via deux chemins : le chemin direct et le chemin décomposé (i.e. par les constituants du mot complexe). Ces deux chemins sont en compétition et ce qui détermine celui qui sera emprunté est l'activation au repos, qui est influencée notamment par la fréquence d'utilisation des éléments considérés. Plus un

élément a une activation au repos élevée, plus il sera facilement disponible dans le lexique. Ainsi, Hay défend l'idée que ce qui est pertinent dans l'accès lexical des mots complexes, ce n'est pas simplement la fréquence des dérivés, mais également la fréquence de leur base puisqu'une base plus fréquente que le dérivé favorisera l'accès par le chemin décomposé pour ce dérivé.

Le premier phénomène pour lequel j'ai évalué la pertinence de cette hypothèse est ce que j'ai appelé la préservation accentuelle exceptionnelle : les cas où la Préservation Faible aboutit à une séquence de deux accents adjacents (e.g. *depártment* ♦ *depàrtméntal*). Dans ma thèse et un article qui a exploré ce phénomène plus en profondeur (Dabouis 2019), j'ai pu montrer à l'aide d'un inventaire dictionnaire large que la segmentabilité joue un rôle et interagit avec des contraintes phonologiques puisque cette préservation accentuelle est plus probable si :

- la première syllabe est ouverte ;
- la deuxième syllabe est fermée ;
- la base est plus fréquente que le dérivé.

Dans l'article de 2019, je développe une analyse à l'aide d'un modèle MaxEnt, qui est une version probabiliste d'OT qui permet de dériver des probabilités à partir des infractions aux différentes contraintes. J'inclus également la segmentabilité comme une probabilité – externe à la phonologie – que le dérivé soit généralement récupéré dans le lexique par la voie directe ou la voie décomposée. Par ailleurs, mes résultats ont montré que la segmentabilité des suffixes concernés pourrait jouer un rôle, puisque les suffixes trouvés dans des mots généralement plus segmentables étaient plus souvent trouvés dans des dérivés avec préservation accentuelle exceptionnelle.

Je me suis également intéressé à la Préservation Faible lorsque celle-ci ne génère pas de suites d'accents adjacents, mais plutôt lorsqu'elle échoue (e.g. *antícipate* ♦ *ànticipátion* ~ *antícipátion*).¹⁶¹ Collie (2007, 2008) a montré que ces échecs de préservation sont liés à la segmentabilité des dérivés, mesurée par la fréquence relative des bases et des dérivés. Elle se basait sur des données dictionnaires, mais les mots retenus présentaient plusieurs problèmes et il semblait utile de répliquer l'étude. C'est donc ce que nous avons fait avec Sabine Arndt-Lappe dans un article en cours de révision (Arndt-Lappe & Dabouis soumis). Dans cet article, nous étudions la Préservation Faible dans des dérivés suffixaux ayant trois syllabes prétoniques et dont la base locale a un accent sur sa deuxième syllabe, à l'aide de données dictionnaires et de données orales d'une étude d'élicitation. Dans les deux cas, nous trouvons des effets du poids relatif des deux syllabes et de la préfixation opaque, bien que les deux ne puissent être testés en même temps dans les données orales. Nous ne trouvons des effets de la segmentabilité que dans les données orales, et de deux types :

- des effets de la fréquence de la base morphologique, comme rapporté par Collie ;

¹⁶¹ Mes travaux de thèse, également synthétisés dans un article (Dabouis 2020b), ont évalué la prévalence de la Préservation Faible de manière générale. Celle-ci est très largement le comportement majoritaire, et j'ai pu montrer que les échecs de préservation évoqués ici se produisent toujours lorsque la base est accentuée sur sa deuxième syllabe, jamais lorsqu'elle l'est sur sa première syllabe. J'ai interprété cette observation comme un indice du fait que l'accent secondaire est par défaut placé sur la première syllabe, ce qui est cohérent avec le comportement majoritaire observé dans les mots sans structure interne (voir §2.3).

- des effets de la fréquence et de la transparence sémantique de ce que nous appelons la « base nominale », qui est le produit d’une option alternative de segmentation dans les mots à la fois préfixés et suffixés. Par exemple, la base morphologique dans *irrationality* est *irrational*, et la base nominale est *rationality*.

Les différences observées entre ces deux sous-études sont probablement dues à des différences dans le profil fréquentiel des dérivés étudiés, qui n’étaient pas les mêmes dans les deux cas, et peut-être dans la nature des données (dictionnairiques vs. orales).

La troisième étude dans laquelle nous nous sommes intéressés aux éventuels effets de segmentabilité est l’étude sur la réduction vocalique menée avec Jean-Michel Fournier et déjà évoquée au §2.2. Les résultats portant sur les mots non-dérivés servant de point de référence, nous nous sommes intéressés aux effets de l’existence d’une base accentuée sur la syllabe étudiée dans les dérivés à déplacement accentuel (e.g. *vítal* ♦ *vitálicity*, *infórm* ♦ *informátion*). Nous évaluons ainsi le rôle de la fréquence relative et de la transparence sémantique entre la base et le dérivé à l’aide d’un inventaire d’environ 750 dérivés suffixaux tirés de Wells (2008) et Jones (2006). Dans cet inventaire, nous confirmons les effets de la structure syllabique, de la fréquence du mot et de la préfixation opaque déjà observés dans les mots non-dérivés. Quant aux effets de la segmentabilité, ils sont incohérents : on trouve un effet de transparence sémantique, les dérivés sémantiquement transparents sont plus susceptibles d’avoir une voyelle pleine que ceux qui ne le sont pas, mais on ne trouve un effet de la fréquence de la base que dans la position initiale prétonique, et dans la direction inverse de celle prédite par l’hypothèse de la segmentabilité : plus la base est fréquente, plus on observe de voyelles réduites. Ce résultat est cohérent avec les résultats d’une étude précédente de Hammond (2003) sur la position intertonique.

A la suite d’études ayant cherché à tester les effets de la segmentabilité sur la réalisation phonétique de séquences de consonnes nasales identiques appartenant à un préfixe et à sa base (e.g. *immoral*, *unnatural* ; Oh & Redford 2012 ; Oh 2013 ; Ben Hedia & Plag 2017 ; Yu 2022), nous avons entrepris avec Olivier Glain et Sylvain Navarro d’étudier le même phénomène pour /ɪ/ (Dabouis *et al.* soumis). En plus de répliquer les études menées sur les nasales, nous souhaitons tester une étude utilisant des données quasi-spontanées et établir s’il existait des différences entre anglais britannique et américain. Nous avons donc étudié la réalisation de 25 mots commençant par la séquence <irr-> dans des extraits audios tirés de YouGlish, une plateforme permettant d’avoir accès à des prononciations de mots ou expressions dans des vidéos YouTube. Nous avons récupéré 500 occurrences pour ces 25 mots en anglais britannique et 1046 en anglais américain (ce dernier étant plus représenté sur YouGlish). Nous avons alors mesuré la durée du [ɪ] et du [ɪ] sous Praat (Boersma & Weenik 2024) et avons étudié les effets de différentes variables sur cette durée : présence d’un accent sur la deuxième syllabe, transparence sémantique, genre, variété d’anglais, fréquence de la base et du dérivé. Les effets trouvés vont globalement dans le même sens que les études précédentes, de sorte que [ɪ] est plus long si :

- la vitesse d’élocution est plus lente ;
- la fréquence du mot est plus basse ;
- le mot est sémantiquement transparent ;
- la deuxième syllabe du mot est accentuée ;

- le locuteur est un homme ;
- le locuteur a pour variété l'anglais américain.

En revanche, les 19 mots formés à partir d'une base libre et du préfixe *ir-* révèlent des effets plus surprenants de la fréquence relative et de la fréquence de la base. En effet, quelle que soit la manière dont la fréquence de la base est incorporée dans le modèle statistique, on trouve que plus la fréquence de la base est élevée, plus le [ɪ] est court. Ce résultat contredit les prédictions de l'hypothèse de la segmentabilité (qui prédirait un [ɪ] plus long, la frontière étant plus « forte »), là où la transparence sémantique la confirme. Une possibilité est que cet effet n'est peut-être pas un effet de la segmentabilité mais un effet de réduction induit par la fréquence dans la base et reproduit dans le dérivé.

Ainsi, à la suite d'études menées par d'autres chercheurs, les études auxquelles j'ai participé et qui cherchaient à tester l'hypothèse de la segmentabilité ont des résultats mitigés et parfois incohérents. Ces résultats sont une des raisons d'explorer d'autres modèles, que l'on évoquera dans la section suivante.

3.6. Au-delà de la localité et de la contenance

Cela fait plus d'une trentaine d'années que différentes théories phonologiques sont en désaccord au sujet des dépendances paradigmatiques et en particulier l'identification des membres de la famille phonologique d'un dérivé pouvant agir comme base phonologique pour ce dérivé.

Un premier point de désaccord porte sur la **localité**, i.e. le fait que seule la base locale serait une base phonologique possible, ce qui est la position habituelle des modèles cycliques. Cependant, des études auxquelles j'ai participé montrent des effets des bases distantes (e.g. *collect* ♦ *collective* ♦ *collectivity*) dans la Préservation Faible. Dans mon article de 2019 sur la préservation accentuelle exceptionnelle, je montre que prendre la fréquence de la base la plus fréquente entre la base locale et la base distante améliore significativement les modèles statistiques, ce qui peut être interprété comme un indice que la base distante peut servir de base phonologique. On trouve des effets comparables dans notre article sur la Préservation Faible (Arndt-Lappe & Dabouis soumis) lorsque la base distante est accentuée comme la base locale (*sustáin* ♦ *sustáinable* ♦ *sustainability*), mais pas lorsque les deux bases sont accentuées différemment (e.g. *súbstance* ♦ *substántiate* ♦ *substantiátion*).

Comme nous l'avons vu dans la section précédente, on trouve des effets de fréquence pour **différentes bases locales possibles** (e.g. *impossibility*: *im-* + *possibility* ou *impossible* + *-ity*) dans l'étude de la Préservation Faible également. Des effets similaires sont rapportés par Arndt-Lappe & Ben Hedia (2019) dans l'accentuation des adjectifs en *-atory*, qui semblent influencés par leur base verbale comme par leur co-dérivé en *-ation*. Ceci suggère qu'ils peuvent être analysés comme une juxtaposition de *-(at)ory* à une base verbale (e.g. *oscillatory* = *oscillate* + *-ory*) ou comme une juxtaposition de *-ory* à une base nominale en *-ation* avec troncation de *-ion* (e.g. *oscillatory* = *oscillat(ion)* + *-ory*).

Enfin, un débat central oppose les tenants de OO-Correspondence et ceux des modèles cycliques : la base phonologique peut-elle être un membre de la famille morphologique du dérivé qui n'est **pas contenue** dans celui-ci ? Les défenseurs d'OO-Correspondence, qui

rendent compte des cas de dépendance paradigmatique par des contraintes de correspondance entre formes de surface plutôt que par le cycle, défendent que cela est indispensable pour analyser le « conservatisme lexical ». Ce phénomène désigne des préférences grammaticales contre l'utilisation de formes qui ne s'appuient pas sur des précédents lexicaux (Steriade 1997, 1999). Par exemple, l'adjectif *remédiable* diffère accentuellement de sa base locale *remedy*, mais pas *párodiable* de sa base *párody*. Steriade défend que cette différence soit due à des différences dans la famille morphologique de ces deux mots : l'accentuation de *remédiable* peut se baser sur *remédial*, mais on n'a pas de **paródial* qui permettrait un éventuel **paródiabile*. Ceci ouvre la porte à la possibilité que de multiples bases phonologiques soient disponibles et pose la question des facteurs guidant le choix d'une base plutôt qu'une autre ou, si l'on suppose que plusieurs bases peuvent agir comme base phonologiques simultanément, la question de l'importance accordée à chaque base. Les trois propositions qu'on trouve dans la littérature sont **l'activation au repos** (influencée par la fréquence notamment), **la marque** (des bases non-locales seraient sélectionnées si elles améliorent le caractère non-marqué du dérivé) et **la priorité à la base locale**. Ces facteurs sont à ce jour peu étudiés, et il semble que les conditions dans lesquelles des dépendances paradigmatiques au-delà de la base locale puissent émerger sont peu courantes, ce qui expliquerait l'efficacité de modèles théoriques ne reconnaissant que la base locale comme base phonologique possible.

4. ORTHOGRAPHE ET PHONOLOGIE

La relation entre orthographe et phonologie fait l'objet de nombreuses études par les psycholinguistes et les spécialistes d'acquisition de la lecture, mais peu par les phonologues. Ainsi, les travaux en phonologie qui évoquent la question de l'orthographe se font rares. Les rares qui le font invoquent l'orthographe pour rendre compte de phénomènes pour lesquels aucune explication phonologique ne semble satisfaisante. L'Ecole de Guierre s'intéressant aux relations écrit-oral depuis ses débuts, j'ai travaillé sur cette question tant sur le plan plus théorique (ce qui a abouti à deux articles, Dabouis 2014 et Dabouis 2023a, et plusieurs communications) que sur le plan empirique, puisque j'ai étudié les correspondances écrit-oral dans les voyelles accentuées dans mon mémoire et dans ma thèse.

Mes lectures en psycholinguistique, notamment sur l'acquisition de la lecture m'ont appris que l'apprentissage de **l'orthographe est une force considérable de restructuration des connaissances phonologiques**, qui réorganise de manière permanente le système phonologique. Ses effets sont aussi particulièrement visibles dans l'adaptation des emprunts et dans l'acquisition d'une langue seconde. Je défends donc l'idée, s'agissant de l'anglais, d'étudier le système phonologique en tenant compte de l'orthographe. Il ne s'agit peut-être pas de phonologie « pure », mais de quels usagers de l'anglais étudierions-nous le système phonologique si nous supposons que l'orthographe ne doit y avoir aucun rôle ? Une part considérable de la population anglophone native sachant lire et écrire, il semble plus souhaitable d'étudier le système phonologique de locuteurs ayant accès à l'information orthographique.

Ainsi, j'ai répertorié l'ensemble des phénomènes que j'ai pu identifier pour lesquels le recours à l'information orthographique semble nécessaire car elles dépassent les limites de ce que la phonologie peut faire, notamment des alternances vocaliques (e.g. *divine* /di'vajn/ ♦

divinity /di'vɪnəti/, *brutal* /'brʊwtəl/ ♦ *brutality* /brʊw'tæləti/) et consonantiques (e.g. *electric*/k/ ♦ *electric/s/ity*, *alle/g/ation* ♦ *alle/dʒ/e*, *sign* /'sɑjn/ ♦ *signature* /'sɪɡnətʃə/) qui sont soit jugées « non-naturelles » soit pour lesquelles l'information nécessaire pour former le dérivé à partir de la base se trouve uniquement dans la graphie. J'ai également exploré quelques options théoriques cherchant à répondre à la question de la nature de la restructuration des connaissances phonologiques induites par l'apprentissage de l'orthographe :

- Création de connexions entre voyelles phonologiquement distantes (e.g. <i></i> - /{ɪ, əj, ə:, əjə}/) ;
- Insertion de segments additionnels dans la représentation phonologique (e.g. *damn* /'dam/ → /'damn/, *family* /'famlɪj/ → /'faməlɪj/) ;
- Désambiguïsation de traits généralement neutralisés en surface (e.g. /t/ vs. /d/ dans *write* ['ɹaɪt] ~ *writer* ['ɹaɪtə]; *ride* ['ɹaɪd] ~ *rider* ['ɹaɪtə]) ;
- Création d'un niveau hybride « graphophonologique » ;
- Processus phonologiques faisant seulement occasionnellement référence à l'orthographe.¹⁶²

Une option pourrait aussi être de travailler dans le cadre du modèle BiPhon, pour lequel une extension intégrant l'orthographe a été proposée par Hamann & Colombo (2017). Malgré mes tentatives de développer un système complet pour une grammaire de lecture intégrant les correspondances écrit-oral pour les voyelles ainsi que l'accentuation, beaucoup de travail reste à faire pour qu'un tel système soit satisfaisant.

Sur le plan des correspondances écrit-oral pour les voyelles accentuées, l'École de Guierre avait montré que, globalement, les voyelles avec un accent secondaire prétonique suivaient les mêmes règles que les voyelles avec un accent primaire, mais des questions subsistaient sur les règles relatives aux voyelles monographe autres que <u> suivies d'une seule consonne. Dans ce contexte, les règles proposées divergent selon l'analyse qui est faite de la séquence prétonique, selon qu'on inclut la syllabe portant l'accent primaire et celles qui la suivent dans le contexte déterminant des règles ou non. J'ai donc entrepris d'analyser quantitativement ces règles dans ma thèse et dans un article paru en 2018 et j'ai trouvé que :

- le comportement largement majoritaire lorsque la voyelle avec accent secondaire prétonique est suivie d'une syllabe inaccentuée est d'avoir une voyelle entravée (brève, e.g. *vòluntéer*, *màcaróon*, *rècomménd*) ;
- lorsque la voyelle est initiale et prétonique, elle est libre (une diphtongue), ce qui n'avait pas été relevé dans les études précédentes (e.g. *bètíse*, *dèméntia*, *hòtél*, *migráte*) ;
- dans les suffixés à déplacement accentuel, la voyelle est systématiquement reprise de la base lorsque celle-ci contient une voyelle pleine (e.g. *neg/əw/tiátion* ♦ *neg/əw/tiate*, *fam/ɪ/liárity* ♦ *fam/ɪ/liar*).

¹⁶² Il est probable qu'une personne graphématisée intègre également une contrainte interdisant la réalisation de sons n'ayant pas de contrepartie graphique, et que cette contrainte soit fortement sujette à un conditionnement sociolinguistique. Une telle contrainte pourrait expliquer le fait que des sons ayant une représentation graphique (e.g. [p] dans *Sampson*, [ɪ] dans *door outside*) soient réalisés plus fréquemment que des sons non représentés dans la graphie (e.g. [p] dans *Samson*, [ɪ] dans *salsa-evening*).

Il est intéressant de noter que, dans les suffixés à déplacement accentuel, les seuls cas où la voyelle n'est pas reprise concernent des cas où elle est libre dans le dérivant, jamais entravée, ce qui suggère que la valeur entravée est la valeur par défaut dans ce contexte.

Dans mon article de 2018 sur ce sujet, je propose une analyse des schémas observés en termes de structure en pieds minimalement bimoraïques. J'analyse alors la prédominance de voyelles entravées lorsque la syllabe avec un accent secondaire est suivie d'une syllabe inaccentuée comme un cas d'abrègement trochaïque, comme proposé par Prince (1991) pour l'accent primaire. Dans cet article, je propose également que cette analyse puisse être étendue à l'analyse des différences de distributions des voyelles trouvées entre les verbes et les noms. Depuis, j'ai pris connaissance des travaux de Nessly (1982) et Treiman & Kessler (2023), qui montrent que certaines terminaisons sont fortement associées à des valeurs de voyelle dans la syllabe précédente (e.g. <-it> est particulièrement associé aux voyelles entravées, e.g. *crédit*, *digit*, *limit*), et que ces terminaisons sont distribuées inégalement entre verbes et noms. Par conséquent, les différences observées entre ces deux catégories syntaxiques pourraient être liées à des différences de terminaisons, et des recherches supplémentaires sont nécessaires pour clarifier ce point.

5. LES SOUS-LEXIQUES DE L'ANGLAIS

L'idée que différentes classes de mots (et notamment les emprunts) suivent des règles phonologiques différentes est assez répandue. Cependant, il est courant de trouver des mots d'emprunt comme exemples illustrant les règles phonologiques de l'anglais (e.g. *Ticonderoga*, *Monongahela*), ce qui suggère que tous les phonologues travaillant sur l'anglais ne font pas cette distinction. Dans l'approche guierrienne, il est courant de distinguer les parties natives et étrangères du vocabulaire, ce qui peut s'expliquer par l'attention portée aux relations écrit-oral, où ces distinctions sont indispensables (comme le note Carney 1994). J.-M. Fournier (2007) va plus loin en affirmant que l'anglais, et particulièrement son système accentuel, est un système hybride ayant hérité non pas une mais plusieurs phonologies de ses origines germaniques et romanes.

Dans mes études de la réduction vocalique ou de l'accent secondaire prétonique, j'ai donc inclus une variable relative au caractère étranger des mots pour évaluer ses effets. Par la suite, avec Pierre Fournier, nous avons entrepris de formaliser et systématiser cette analyse et avons proposé un modèle des différentes phonologies de l'anglais, correspondant à différents sous-lexiques, dans un chapitre paru en 2022 (Dabouis & P. Fournier 2022). Nous y proposons une distinction entre quatre sous-lexiques principaux : §CORE, §FRENCH, §FOREIGN et §LEARNED. Nous nous sommes appuyés pour cela sur de nombreuses études précédentes que nous avons croisées afin d'identifier les caractéristiques de ces sous-lexiques, sachant que certaines de ces caractéristiques peuvent être partagées par plusieurs sous-lexiques (e.g. la correspondance <a>/a:/ se trouve aussi bien dans §FRENCH que §FOREIGN). Nous supposons que ces sous-lexiques sont en relation indirecte avec les langues sources, de sorte que des différences avec les langues sources peuvent émerger si les mots concernés présentent des caractéristiques qui diffèrent de celles généralement trouvées dans le sous-lexique correspondant (e.g. si un mot italien a un accent antépénultième, on pourrait s'attendre à une

adaptation vers l'accent pénultième en anglais car il est probable qu'il soit classifié dans §FOREIGN, où l'accent pénultième est prédominant). Nous avons dérivé une série de prédictions de ce modèle qui devront être testées dans des recherches futures.

Dans un article paru en 2024 (Dabouis & P. Fournier 2024), nous explorons plus en détail les propriétés accentuelles des emprunts au français en anglais britannique et américain sur la base d'un inventaire de 1043 entrées tirées de Wells (2008). Nous trouvons que la prévalence de l'accent primaire final, supposé typique des emprunts au français, varie en fonction de :

- la complexité (définie ici comme la présence d'un espace ou d'un tiret graphique), celle-ci favorisant l'accent final ;
- le nombre de syllabes, l'accent final étant plus courant dans les entrées plus courtes ;
- le poids des syllabes, l'accent final étant plus courant dans les syllabes plus lourdes.

Nous avons également trouvé que certaines terminaisons favorisaient l'accent final et que d'autres le défavorisaient, et que cela variait légèrement entre anglais britannique et américain. Par ailleurs, un résultat que l'on retrouve dans tous les sous-inventaires étudiés est la plus grande prévalence de l'accent primaire final en anglais américain qu'en anglais britannique. En effet, dans l'inventaire complet, 78% des entrées ont l'accent primaire en position finale en anglais américain, contre seulement 50% pour l'anglais britannique. Nous avons également trouvé que, dans les entrées simples, lorsque l'accent primaire n'est pas sur la finale, il est généralement sur l'antépénultième. Par ailleurs, nous avons noté la forte présence de voyelles pleines (analysables comme un accent secondaire post-tonique) en position finale lorsque la dernière syllabe ne porte pas l'accent primaire. Ceci peut être analysé comme une préférence du sous-lexique §FRENCH pour un pied monosyllabique aligné à droite, qu'il s'agisse de la tête du mot phonologique (= accent primaire) ou non (= accent secondaire).

Enfin, nous avons exploré ces sous-systèmes de manière plus extensive avec Marie Gabillet dans le cadre du projet ERSaF en réanalysant des données tirées de Moore-Cantwell (2020). Dans cet article, Moore-Cantwell affirme que l'accentuation des trisyllabes est influencée par la nature du noyau de la syllabe finale, la voyelle /ij/ (qu'elle transcrit /i/) favorisant l'accentuation antépénultième. Notre hypothèse est que l'asymétrie entre différents noyaux est attribuable à une asymétrie dans la distribution des différents noyaux entre les différents sous-lexiques. Autrement dit, si l'on trouve plus d'accentuation antépénultième pour /ij/, c'est que ce noyau est surreprésenté dans des sous-lexiques où l'accentuation antépénultième est plus courante. Nous avons donc réanalysé ses données dictionnaires tirées du dictionnaire de Carnegie Mellon University (CMU). Nous avons reclassifié certaines données sur le plan morphologique, la classification ayant été faite automatiquement, et avons annoté les mots comme appartenant à l'un des quatre sous-lexiques proposés par Dabouis & P. Fournier (2022) sur la base de correspondances écrit-oral et de terminaisons qu'ils répertorient. Notre nouvelle analyse des données après la reclassification morphologique ne fait pas apparaître de comportement spécifique de /ij/, mais plutôt de /ə/, qui est plus souvent associé à l'accentuation pénultième que les autres noyaux étudiés. Nous avons également mené différentes analyses statistiques (régressions logistiques et arbres d'inférence conditionnelle), qui montrent que les sous-lexiques sont le principal facteur déterminant, distinguant les mots de §FOREIGN de ceux de §CORE et §LEARNED (ceux de §FRENCH ayant été exclus des analyses

car trop peu nombreux). On trouve un petit effet des noyaux au sein de §CORE et §LEARNED, mais les effets sont soit marginalement significatifs, soit trouvés dans peu de mots. Enfin, nous avons proposé une explication alternative pour les résultats expérimentaux de Moore-Cantwell, qui montrent une différence entre des non-mots finissant en /ə/ et des non-mots finissant en /ɪj/, les derniers étant produits plus souvent avec un accent antépénultième que les premiers. Il est possible que cette différence soit due au fait que les participants aient reproduit la distribution de ces noyaux finaux entre §FOREIGN et §CORE trouvée dans le lexique et aient assigné les non-mots en /ɪj/ plus souvent à §CORE (environ 2/3 mots comparables au non-mots utilisés dans le CMU ont été classifiés comme §CORE) que /ə/, noyau final très majoritairement trouvé dans §FOREIGN dans les données du CMU. Cependant, l'étude expérimentale de Moore-Cantwell n'était pas conçue pour tester ce modèle, et les recherches futures devraient tester directement ce modèle pour établir sa validité.

6. CONCLUSION

Cette synthèse de mes travaux m'a permis de réaliser mon évolution sur les treize dernières années. Si les deux caractéristiques de mes travaux (recours systématique aux données et diversité des approches théoriques) sont restées constantes, je suis passé des préoccupations de l'Ecole de Guierre à une diversité de sujets liés à la structure prosodique, au poids des syllabes et à l'organisation du lexique. Si je devais identifier les aspects qui rassemblent les sujets relatifs à la phonologie des mots en anglais auxquels je me suis intéressé, j'en proposerais deux. Le premier est mon intérêt pour les **représentations** : la nature de l'accent et sa relation aux accents mélodiques, de la manière de représenter les différents comportements phonologiques des suffixes, de la connaissance sur les mots dont les formants sont sémantiquement appauvris et de la relation que l'orthographe entretient avec la phonologie. Le deuxième est mon intérêt pour les **interfaces** : une part considérable de mes travaux implique une interaction de la phonologie avec la morphosyntaxe, mais traite aussi de son interaction avec l'orthographe et l'organisation du lexique. A l'avenir, je souhaite poursuivre certaines de ces thématiques, notamment les corrélats segmentaux de l'accentuation et des différentes classes de suffixes. Au sein du projet ERSaF, nous allons continuer à explorer les caractéristiques des préfixés opaques, le rôle des sous-lexiques et comment des modèles computationnels peuvent rendre compte de leurs propriétés. J'aimerais également explorer les prédictions faites par le modèle développé avec Pierre Fournier sur les sous-lexiques.

Sur le plan des méthodes empiriques, mon évolution a été de passer de simples statistiques descriptives aux statistiques inférentielles, des dictionnaires de prononciations aux expérimentations et aux données orales. Je continue en approfondissant mes connaissances des statistiques inférentielles et me forme aux statistiques bayésiennes. Je souhaite également me former plus amplement aux modèles computationnels comme AML et NDL et à ce qu'on peut en apprendre, et apprendre plus sur l'analyse acoustique.

Enfin, j'ai été formé dans l'Ecole de Guierre et ai rapidement étudié d'autres modèles théoriques comme celui de *SPE*, la Phonologie Métrique et la Phonologie Lexicale. L'éventail de théories qui me sont familières est maintenant bien plus étendu, que ce soient les théories phonologiques, morphologiques ou psycholinguistiques. La préparation de ce document y a

contribué puisqu'elle a été l'occasion pour moi de me renseigner sur des modèles que je maîtrisais moins. Loin d'embrasser une théorie en particulier, je souhaite à l'avenir continuer à apprendre des différents modèles existants, et de mettre au jour ce que la phonologie de l'anglais peut leur apporter.


CURRICULUM VITAE


Quentin Dabouis

Maître de conférences
Université Clermont Auvergne
Laboratoire de Recherche sur le Langage (UR 999)
Membre associé du LLL (UMR 7270)

 <http://lrl.uca.fr/rubrique166.html>

 quentin.dabouis@uca.fr

 06 79 74 86 95

 13 rue du parc de Montjuzet
63100 Clermont-Ferrand



HAL
science ouverte



FORMATION ET DIPLOMES

- 2012 – 2016** **Doctorat en linguistique anglaise** - *Université de Tours*
Intitulé de la thèse : « L'accent secondaire en anglais contemporain »
Sous la direction de Jean-Michel Fournier et Véronique Abasq
Membres du jury : Ricardo Bermúdez-Otero, Sophie Herment, Sylvie Hanote et Ives Trevian

Laboratoire Ligérien de Linguistique (UMR 7270)
Qualifié en 7^{ème} et 11^{ème} sections CNU
- 2010 – 2012** **Master de linguistique avancée et applications** - *Université de Tours*
Mention très bien.
Intitulé du mémoire : « Accentuation et contrainte en anglais contemporain : Le cas des adjectifs suffixés en *-al* »
Sous la direction de Jean-Michel Fournier et Véronique Abasq
- 2007 – 2010** **Licence LLCE spécialité anglais** - *Université de Tours*

POSTES DANS L'ENSEIGNEMENT SUPERIEUR ET LA RECHERCHE

- SEPT 2018 – PRÉSENT** **Maître de Conférences** *Université Clermont Auvergne*
- SEPT 2017 – JUIN 2018** **Enseignant contractuel** *Université de Tours*
- SEPT 2015 – AOUT 2017** **A.T.E.R.** *Université de Tours*
- SEPT 2013 – JUIN 2015** **Enseignant contractuel d'anglais** *École Polytechnique de l'Université de Tours*

RESPONSABILITES COLLECTIVES

Université Clermont Auvergne

- 2024 - PRÉSENT** **Responsable pédagogique de la 2^{ème} année** de licence LLCER option anglais
- Organisation de la réunion de rentrée
 - Diffusion des informations importantes tout au long de l'année
 - Rencontre avec les étudiants à leur demande
 - Gestion de la répartition dans les groupes de TD
 - Rédaction et mise en page du livret de l'étudiant
 - Suivi des Modalités de Contrôle des Connaissances et des Compétences et leur mise à jour
 - Organisation des jurys de fin de semestre
 - Echanges réguliers avec la secrétaire du département sur l'inscription pédagogique des étudiants, l'organisation des examens et la gestion des notes
 - Organisation de l'élection des délégués de promotion et échanges réguliers avec ces derniers tout au long de l'année
- 2019 – PRÉSENT** **Responsable des lecteurs entrants**
- Recrutement
 - Accueil sur le site et dans le département
 - Accompagnement dans les démarches administratives, l'évaluation et les interactions avec le secrétariat
- 2020 – PRÉSENT** **Membre du conseil du LRL et responsable du séminaire mensuel du LRL.**
- Participation aux décisions du conseil du laboratoire, notamment sur les dépenses du laboratoire et les profils de postes
 - Organisation du séminaire : recherche des intervenants, diffusion de sondages pour fixer les créneaux, réservation de salles, diffusion de l'information auprès des collègues et des étudiants
- 2020 – PRÉSENT** **Professeur-évaluateur dans le cadre de FEI** (mobilités sortantes d'assistants de langue)
Entretiens d'évaluation en vue d'une mobilité sortante
- 2021 – PRÉSENT** **Référent développement durable** du département d'anglais et du LRL et membre du comité de pilotage du réseau des référents développement durable de l'UCA (2020-2024).
- Participation à la création du réseau : mise en place du fonctionnement, rédaction de la lettre de mission
 - Réunions bimensuelles avec le comité de pilotage et la Vice-Présidente à la transition écologique et sociétale
 - Réunions avec le Président de l'UCA deux à trois fois par an
 - Organisation des assemblées générales trois fois par an
 - Accueil des nouveaux référents
 - Organisation d'ateliers de Fresque du Climat pour les référents, les collègues et les étudiants
 - Réalisation du bilan carbone du LRL
 - Diffusion d'informations relatives aux activités du réseau

- 2019 – 2023** **Responsable pédagogique de la 3^{ème} année** de licence LLCER option anglais
- Organisation de la réunion de rentrée
 - Diffusion des informations importantes tout au long de l'année
 - Rencontre avec les étudiants à leur demande
 - Gestion de la répartition dans les groupes de TD
 - Rédaction et mise en page du livret de l'étudiant
 - Suivi des Modalités de Contrôle des Connaissances et des Compétences et leur mise à jour
 - Organisation des jurys de fin de semestre
 - Echanges réguliers avec la secrétaire du département sur l'inscription pédagogique des étudiants, l'organisation des examens et la gestion des notes
 - Organisation de la réunion d'information sur les masters proposés par le département
 - Organisation de l'élection des délégués de promotion et échanges réguliers avec ces derniers tout au long de l'année

Université de Tours (Polytech)

- 2014 – 2015** **Responsable d'année** pour la 3^{ème} année en Electronique et Mécanique et 3^{ème} et 5^{ème} année en Informatique Industrielle.
- Coordination pédagogique de l'équipe d'enseignants d'anglais
 - Organisation des examens
- 2014 – 2015** **Mise en place du CRL** (Centre de Ressources en Langues)
- Sélection des ressources
 - Aménagement
 - Mise en place du fonctionnement du CRL
- 2014 – 2015** Jury pour le concours CLES 2 (en janvier 2014 et 2015).

COMITES DE SELECTION ET COMMISSIONS AD HOC

- 2024** Poste de MCF en linguistique anglaise, Université Clermont Auvergne (président du comité), poste MCF 0827.
- 2024** Poste de MCF en linguistique anglaise, Université Grenoble Alpes, poste MCF 1341.
- 2023** Poste de PRAG en anglais, Polytech Clermont, poste SD0940-G127.
- 2023** Poste de MCF en linguistique et phonologie, Université Jean Moulin Lyon 3, poste MCF 0728.
- 2022** Poste de PRAG en anglais, Polytech Clermont, poste SD0789-51.
- 2022** Poste de PRAG en littérature anglo-américaine, Université Clermont Auvergne, poste SD1102-G290.

Participation à trois jurys de recrutement d'enseignants contractuels d'anglais à l'Université de Tours (2021, 2022, 2024).

AUTRES RESPONSABILITES

2022 Président de jury de baccalauréat, Lycée des métiers Henri Sainte-Claire Deville, Issoire

FORMATIONS SUIVIES

2024 Comprendre comment les étudiant.e.s apprennent pour améliorer ses pratiques d'enseignement
2018 Innover dans son enseignement pour favoriser l'apprentissage en profondeur
2018 Rendre les étudiants actifs

ENSEIGNEMENT

Université Clermont Auvergne – MCF

	L1	L2	L3	M1	M2	Concours	Autre
2023-2024 201,1h		Phonétique & phonologie	Variations phonétiques et phonologiques	Séminaire de linguistique anglaise (mutualisé M1-M2)		AGREG. Phonologie	
			Initiation à la recherche en linguistique	MEEF Phonologie	MEEF Phonologie		
2022-2023 196,8h		Phonétique & phonologie	Variations phonétiques et phonologiques	Séminaire de linguistique anglaise (mutualisé M1-M2)		AGREG. Phonologie	Module dév. durable (2h en ligne)
			Initiation à la recherche en linguistique	MEEF Phonologie	MEEF Phonologie		Module « Sensibiliser au dérèglement climatique » (collège des écoles doctorales)
2021-2022 250,1h	Phonétique & phonologie	Phonétique & phonologie	Variations phonétiques et phonologiques	Séminaire de linguistique anglaise (mutualisé M1-M2)		AGREG. Phonologie	
			Initiation à la recherche en linguistique	MEEF Phonologie	MEEF Phonologie		
				Approches théoriques			

2020-2021 203,6h	Phonétique & phonologie	Phonétique & phonologie	Variations phonétiques et phonologiques Initiation à la recherche en linguistique	Séminaire de linguistique anglaise (mutualisé M1-M2) Approches théoriques	AGREG. Phonologie
2019-2020 245,2h	Phonétique & phonologie	Phonétique & phonologie	Morpho-phonologie et variations Grammaire comparée	Séminaire de linguistique anglaise (mutualisé M1-M2) Approches théoriques	AGREG. Phonologie
2018-2019 160h	Phonétique & phonologie	Phonétique & phonologie	Grammaire comparée	Images et textes	AGREG. Phonologie

Les totaux d'heures enseignées donnés en équivalent TD incluent celles dispensées au centre d'enseignement à distance (CEAD) mais pas les heures attribuées pour les responsabilités collectives ou l'encadrement d'étudiants.

Université de Tours – A.T.E.R (2015-2017) puis contractuel (2017-2018)

	L1	L2	L3	M1	M2
2017-2018 325h	LLCER Grammaire	LLCER Linguistique	LLCER Phonologie	LINGUISTIQUE Morpho-phonologie	LINGUISTIQUE Morpho-phonologie
	LLCER Langue orale	DROIT-LANGUE Compréhension/ phonétique			
		SCIENCES DU LANGAGE Anglais			
2016-2017 192h	LLCER Grammaire	LLCER Linguistique	LLCER Phonologie	LINGUISTIQUE Morpho-phonologie	
	LLCER Langue orale	DROIT-LANGUE Compréhension/ phonétique			
	LLCER MOBIL				

2015-2016 192h	LLCER Grammaire	LLCER Linguistique	LLCER Phonologie	LINGUISTIQUE Morpho- phonologie
	LLCER Langue orale	DROIT-LANGUE Compréhension/ phonétique		
	SCIENCES DU LANGAGE Linguistique générale			
	SCIENCES DU LANGAGE Morphologie			

Université de Tours (Polytech) – contractuel d’anglais

	1 ^{ère} année	2 ^{ème} année	3 ^{ème} année	4 ^{ème} année	5 ^{ème} année	Autre
2014-2015 300h	Anglais		Anglais	Anglais	Anglais	M1 LINGUISTIQUE Morpho- phonologie
2013-2014 280h	Anglais		Anglais	Anglais	Anglais	

FORMATION D’ENSEIGNANTS

SEPT 2019 – PRÉSENT	Correcteur de copies d’agrégation en phonologie anglaise <i>CNED</i>
SEPT 2021	Auteur du cours de phonologie anglaise à l’agrégation <i>CNED</i>

PRODUCTIONS SCIENTIFIQUES

Synthèse par type de production et par année

	12	13	14	15	16	17	18	19	20	21	22	23	24	TOTAL
Articles revue int.					1		1	2	2			1	2	9
Chapitres											1	3		4
Actes			1		1									2
Recensions											1			1
Eval. publiques											1			1
Editoriaux													1	1
Com. colloques int.	4	1	4	4	1	6	4	4		3	2	3	3	39
Com. colloques nat.							1				1			2
Com. JE	1		1	1		1	1	1	1	1	2	4	2	16
Com. invit.						1								1
TOTAL	5	1	6	5	3	8	7	7	3	4	8	11	8	76

Articles dans des revues internationales à comité de lecture

Articles soumis

- **DABOUIS, Q.**, GLAIN, O. & NAVARRO, S. (en révision) “Gemination and degemination of /r/ in British and American English”, soumis à *Morphology*.
- ARNDT-LAPPE, S. & **DABOUIS, Q.** (en révision) “Secondary stress, lexical storage, and morphological structure - new evidence from dictionary and speech data”, soumis à *Glossa*.

A paraître

- **DABOUIS, Q.** & FOURNIER, J.-M. (à paraître) “Vowel reduction and preservation in British English”, accepté dans *Phonology*.

2024

- **DABOUIS, Q.** & FOURNIER, J.-M. (2024) “Opaque morphology and phonology: Historical prefixes in English”, *Journal of Linguistics* (publication FirstView). DOI : <https://doi.org/10.1017/S002222672400015X>
- **DABOUIS, Q.** & FOURNIER, P. (2024) “Stress in French Loanwords in British and American English”, *Journal of Linguistics* (publication FirstView). DOI : <https://doi.org/10.1017/S0022226724000136>

2023

- **DABOUIS, Q.** (2023) “Phonological evidence for morphological complexity in English proper names”, *Anglophonia* [En ligne], 36. DOI : <https://doi.org/10.4000/11qbh>

2020

- **DABOUIS, Q.** (2020) “Secondary stress in contemporary British English: An overview”, *Anglophonia* [En ligne], 30. DOI : <https://doi.org/10.4000/anglophonia.3476>
- **DABOUIS, Q.**, ENGUEHARD, G., FOURNIER, J.-M., LAMPITELLI, N. (2020) “The English “Arab Rule” without feet”, *Acta Linguistica Academica.*, 67 (1), 121-134. DOI : <https://doi.org/10.1556/2062.2020.00009>

2019

- ABASQ, V., **DABOUIS, Q.**, FOURNIER, J.-M., GIRARD, I. (2019) “The Core of the English Lexicon: Stress and Graphophonology”, *Anglophonia* [En ligne], 27. DOI : <https://doi.org/10.4000/anglophonia.2317>
- **DABOUIS, Q.** (2019), “When Accent Preservation Leads to Clash”, *English Language and Linguistics*, 23(2), 363-404. URL : <https://hal.science/hal-01548943>

2018

- DABOUIS, Q. (2018) “The pronunciation of vowels with secondary stress in English”, *Corela* [En ligne], 16-2. DOI : <https://doi.org/10.4000/corela.7153>

2016

- DABOUIS, Q. (2016), “Is the adjectival suffix *-al* a strong suffix?”, *Anglophonia* [En ligne], 21. DOI : <https://doi.org/10.4000/anglophonia.754>

Chapitres d’ouvrage

En préparation

- ARNDT-LAPPE, S. & DABOUIS, Q. (en préparation) “Phonology and the Mental Lexicon”, *Theoretical Phonology and the Phonology of English* (P. Honeybone & M. Krämer, eds.), Cambridge: Cambridge University Press.
- DABOUIS, Q. (en préparation) “The Phonology of English”, *International Encyclopedia of Language and Linguistics*, 3rd Edition, (Hilary Nesi & Petar Milin, eds.), Amsterdam & London: Elsevier.

2023

- DABOUIS, Q., FOURNIER, J.-M. (2023) “The Stress Patterns of English Verbs: Syllable Weight and Morphology”, *New Perspectives on English Word Stress* (Ballier, N., Fournier, J.-M., Przewozny, A., Yamada, E., eds.), Edinburgh : Edinburgh University Press, 154-191. URL : <https://shs.hal.science/halshs-04252455>
- DABOUIS, Q. (2023) “English Phonology and the Literate Speaker: Some Implications for Lexical Stress”, *New Perspectives on English Word Stress* (Ballier, N., Fournier, J.-M., Przewozny, A., Yamada, E., eds.), Edinburgh : Edinburgh University Press, 117-153. URL : <https://shs.hal.science/halshs-04252464>
- DABOUIS, Q., FOURNIER, J.-M., FOURNIER, P., MARTIN, M. (2023) “Word Stress and the Guierrian School”, *New Perspectives on English Word Stress* (Ballier, N., Fournier, J.-M., Przewozny, A., Yamada, E., eds.), Edinburgh : Edinburgh University Press, 53-82. URL : <https://shs.hal.science/halshs-04252460>

2022

- DABOUIS, Q. & FOURNIER, P. (2022) “English PhonologieS”, in Arigne V. & Rocq-Migette, C. (eds.), *Modèles et modélisation en linguistique / Models and Modelisation in Linguistics*. Bruxelles, Belgique : Peter Lang Verlag, 215-258. URL : <https://halshs.archives-ouvertes.fr/hal-03783885>

Actes de conférences

2016

- DABOUIS, Q. (2016) “Syllable Weight and Secondary Stress in English Suffixal Derivatives”. *Proceedings of the LLL Young Researchers Conference*, (1), Dabouis, Q., Ndione, A., Ternisien, C. (eds.), 6-23. URL : <https://halshs.archives-ouvertes.fr/hal-01379921v1>

2014

- DABOUIS, Q. (2014) “English Stress and Underlying Representations”, *Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York*, (1), pp. 1-15. URL : <https://halshs.archives-ouvertes.fr/halshs-01386265>

Recensions

2022

- DABOUIS, Q. (2022), John T. Jensen (2022) *The Lexical and Metrical Phonology of English: The Legacy of The Sound Pattern of English*. Cambridge, New York: Cambridge University Press. Pp. xv + 379, *Phonology*, 39(4):725-734. [Lien d'accès.](#)

Evaluations publiques

2022

- **DABOUIS, Q.** (2022), Discussion in: Damulakis, Gean and Andrew Nevins (auths). “An orthographic twist to the Oprah effect”, *Radical: A Journal of Phonology*, 3, 27-28.

Editoriaux / Introductions scientifiques

2024

- **COUPÉ, C., DABOUIS, Q., GLAIN, O. & HUGOU, V.** (2024), “Introduction: The Phonology-Lexicon Interface”, *Lexis: Journal in English Lexicology*, 23. URL : <https://journals.openedition.org/lexis/8123>

Communications dans colloques internationaux avec CL

2024

- **DABOUIS, Q. & GABILLET, M.** (2024), “The effects of final vowels and sublexica on stress in English”, communication réalisée aux 21^{èmes} rencontres du Réseau Français de Phonologie (RFP), à l’Université Jules Verne Picardie, Amiens (26-28 juin).
- **DABOUIS, Q. & GABILLET, M.** (2024), “On stress and accent in English”, communication réalisée au 21^{ème} colloque d’avril sur l’anglais oral (ALOES), à l’Université Sorbonne Paris Nord, France (29-30 mars).
- **GABILLET, M., SEILER, A., ARNDT-LAPPE, S. & DABOUIS, Q.** (2024), “Stress assignment in English pseudo-words: On the role of weight, prefixation and syntactic category”, communication réalisée à la 21^{ème} conférence OCP (Old World Conference in Phonology) à l’Université de Leipzig, Allemagne (14-16 février).

2023

- **DABOUIS, Q.** (2023), “Morphological gradience in phonology”, communication réalisée au 30^{ème} colloque MFM (Manchester Phonology Meeting), à l’Université de Manchester, Royaume-Uni (25-27 mai).
- **DABOUIS, Q., GLAIN, O. & NAVARRO, S.** (2023) “Patterns of /r/ gemination in British and American English: A comparative study”, communication réalisée à la 16^{ème} conférence PAC, à l’Université Paris Nanterre (12-14 avril).
- **DABOUIS, Q.** (2023), “Opaque yet still complex: Proper names in English”, communication réalisée à la 20^{ème} conférence OCP (Old World Conference in Phonology) à l’Université de Tours, France (25-27 janvier).

2022

- **DABOUIS, Q., GLAIN, O. & NAVARRO, S.** (2022), “Does rhoticity entail possible /r/ gemination in English?”, communication réalisée au 20^{ème} colloque d’avril sur l’anglais oral (ALOES), en ligne (1-2 avril).
- **DABOUIS, Q. & FOURNIER, P.** (2022), “Stress placement in French loanwords in English: a dictionary study”, communication réalisée au 20^{ème} colloque d’avril sur l’anglais oral (ALOES), en ligne (1-2 avril).

2021

- **DABOUIS, Q., GLAIN, O. & NAVARRO, S.** (2021) “Towards a reassessment of the gemination of [ɹ] in British and American English?”, poster présenté à la 15^{ème} conférence PAC, en ligne (1-3 septembre).
- **DABOUIS, Q., GLAIN, O., LUC, C. & NAVARRO, S.** (2021) “The gemination of /r/ in British and American English”, communication réalisée aux 18^{èmes} rencontres du Réseau Français de Phonologie (RFP), en ligne (1-2 juillet).
- **DABOUIS, Q. & FOURNIER, P.** (2021) “The multiple phonologies of English”, poster présenté au 28^{ème} colloque MFM (Manchester Phonology Meeting), en ligne (26-28 mai).

2019

- **DABOUIS, Q.**, ENGUEHARD, G. & LAMPITELLI, N. (2019) “Ternary rhythm as a complex morphological domain”, communication réalisée à la conférence RecPhon2019 à l’Université Autonome de Barcelone (21-22 novembre).
- **DABOUIS, Q.** (2019) “Quelles marques de l’autonomie phonologique des préfixes en anglais ?”, communication réalisée aux 17^{èmes} rencontres du Réseau Français de Phonologie (RFP) à l’Université d’Orléans (26-28 juin).
- **DABOUIS, Q.** & FOURNIER, J.-M. (2019) “On pretonic vowel reduction in English”, communication réalisée à la 14^{ème} conférence PAC (Phonologie de l’Anglais Contemporain) au Laboratoire Parole et Langage, Aix-en-Provence (4-5 juin).
- **DABOUIS, Q.** & FOURNIER, J.-M. (2019) “On the role of morphology, syllable structure, frequency and spelling in English vowel reduction”, poster présenté au 27^{ème} colloque MFM (Manchester Phonology Meeting) à l’université de Manchester, Royaume-Uni (23-25 mai).

2018

- **DABOUIS, Q.**, ENGUEHARD, G., FOURNIER, J.-M. & LAMPITELLI, N. (2018) “Towards a footless approach to English vowel reduction” communication réalisée aux 16^{èmes} rencontres du Réseau Français de Phonologie (RFP) au centre CNRS Pouchet à Paris (5-7 juillet).
- **DABOUIS, Q.**, ENGUEHARD, G., FOURNIER, J.-M. & LAMPITELLI, N. (2018) “The English ‘Arab rule’ without metrics”, communication réalisée au 26^{ème} colloque MFM (Manchester Phonology Meeting) à l’université de Manchester, Royaume-Uni (24-26 mai).
- ARNDT-LAPPE, S. & **DABOUIS, Q.** (2018) “Secondary stress and morphological structure – new evidence from dictionary and speech data”, communication réalisée au 18^{ème} International Morphology Meeting à Budapest, Hongrie (10-13 mai).
- ABASQ, V., **DABOUIS, Q.**, FOURNIER, J.-M., GIRARD, I. (2018) “The phonology and morphology of the core of the English lexicon”, communication réalisée au 19^{ème} colloque d’avril sur l’anglais oral (ALOES) à l’Université de Paris 13, France (6-7 avril).

2017

- **DABOUIS, Q.**, ENGUEHARD, G. & LAMPITELLI, N. (2017) “The English ‘Arab rule’: Analysis and Theoretical Implications”, communication réalisée lors de la Government Phonology Round Table (GPRT) à l’université catholique Pázmány Péter à Budapest, Hongrie (18 novembre).
- **DABOUIS, Q.** & FOURNIER, J.-M. (2017) “The Stress Pattern of English Verbs”, communication réalisée à la 13^{ème} conférence PAC (Phonologie de l’Anglais Contemporain) à l’université Paris Nanterre (28-30 septembre).
- **DABOUIS, Q.** & FOURNIER, J.-M. (2017) “L’accentuation des verbes en anglais : Pour un modèle de reconnaissance morphologique”, communication réalisée lors des 15^{èmes} rencontres du Réseau Français de Phonologie (RFP) à l’université de Grenoble (5-7 juillet).
- **DABOUIS, Q.**, FOURNIER, J.-M., GIRARD, I., LAMPITELLI, N. (2017) “Stress in English Long Verbs”, poster présenté au 25^{ème} colloque MFM (Manchester Phonology Meeting) à l’université de Manchester, Royaume-Uni (25-27 mai).
- **DABOUIS, Q.**, FOURNIER, J.-M., GIRARD, I. (2017) “Ternarity is not an issue: Secondary stress is left edge marking”, communication réalisée lors de la 4^{ème} rencontre du groupe de recherche international PTA (Phonological Theory Agora) sur le thème « Ternarity in English » à l’Université de Manchester (24 mai).
- **DABOUIS, Q.** (2017) “Semantically Opaque Prefixes and English Phonology”, communication réalisée à la 14^{ème} conférence OCP (Old World Conference in Phonology) à l’Université Heinrich Heine, Düsseldorf, Allemagne (20-22 février).

2016

- **DABOUIS, Q.** (2016) “When Stress Preservation Leads to Clash”, communication réalisée à la 13^{ème} conférence OCP (Old World Conference in Phonology) à l’Université Eötvös Loránd, Budapest, Hongrie (14-16 janvier).

2015

- **DABOUIS, Q.** (2015), “Spelling, Syllable Weight and English Stress”, communication réalisée lors des 13^{èmes} rencontres du Réseau Français de Phonologie (RFP) à l’Université Bordeaux-Montaigne (29 juin-1^{er} juillet).
- **DABOUIS, Q.** (2015), “Secondary Stress in Contemporary British English”, communication réalisée lors de la 2^{ème} rencontre du groupe de recherche international PTA (Phonological Theory Agora) à Nimègue (10-11 décembre).
- **DABOUIS, Q.** (2015), “English Phonology and the Literate Speaker”, communication réalisée lors du 11^{ème} colloque international PAC (Phonologie de l’Anglais Contemporain) à l’Université de Toulouse II-Le Mirail (9-13 avril).
- **DABOUIS, Q., FOURNIER, J.-M.** (2015), “Dissyllables and Syllable Weight”, communication réalisée lors du 11^{ème} colloque international PAC (Phonologie de l’Anglais Contemporain) à l’Université de Toulouse II-Le Mirail (9-13 avril).

2014

- **DABOUIS, Q.** (2014), “English Stress and Suffixes: Class I or Class II? The Case of the Adjectival Suffix *-al*”, poster présenté au 22^{ème} colloque MFM (Manchester Phonology Meeting) à l’Université de Manchester, Royaume-Uni (21-24 mai).
- **ABASQ, V., DABOUIS, Q., FOURNIER, J.-M., GIRARD, I., MARTIN, M.** (2014), “Les préfixés pluricatégoriels dissyllabiques : Réflexions méthodologiques sur l’analyse de corpus”, Communication réalisée au 17^{ème} colloque d’avril sur l’anglais oral (ALOES) à l’Université de Paris 13, France (4-5 avril).
- **DABOUIS, Q.** (2014), “Accentuation lexicale et suffixation en anglais : Le cas de *-al* adjectival”, Communication réalisée au 17^{ème} colloque d’avril sur l’anglais oral (ALOES) à l’Université de Paris 13, France (4-5 avril).
- **DABOUIS, Q., FOURNIER, J.-M., MARTIN, M.** (2014), “Testing Parameters for Stress Placement: the Case of Dissyllabic Prefixed Verb/Noun Pairs”, poster présenté à la 11^{ème} conférence OCP (Old world Conference in Phonology) à l’Université de Leiden, Pays-Bas (22-25 janvier).

2013

- **DABOUIS, Q.** (2013), “English Stress and Phonological Representations”, communication réalisée lors du colloque PARLAY à l’Université de York, Royaume-Uni (6 septembre).

2012

- **DABOUIS, Q.** (2012), “Types of Derivation and Stress Placement: the Case of Suffixed Adjectives in *-al*”, communication réalisée lors du colloque international « Interfaces in English Linguistics » (IEL) 2012 à l’Université Károli Gáspár à Budapest, Hongrie (12-13 octobre).
- **ABASQ, V., DABOUIS, Q.** (2012), “At the Crossroads Between Morphology, Phonology, Syntax and Semantics: the Prefix *sub-*”, communication réalisée lors du colloque international « Interfaces in English Linguistics » (IEL) 2012 à l’Université Károli Gáspár à Budapest, Hongrie (12-13 octobre).
- **ABASQ, V., DABOUIS, Q., DESCLOUX, E., FOURNIER, J.-M., FOURNIER, P., GIRARD, I., MARTIN, M., VANHOUTTE, S.** (2012), “Stress in Dissyllabic Verb/Noun Pairs”, Poster présenté lors du 20^{ème} colloque international MFM (Manchester Phonology Meeting) à l’Université de Manchester (25-27 mai).
- **ABASQ, V., DABOUIS, Q., DESCLOUX, E., FOURNIER, J.-M., FOURNIER, P., GIRARD, I., MARTIN, M., VANHOUTTE, S.** (2012), “Multicategorical Prefixed Words Stress Behaviour: Variation and Frequency”, communication réalisée lors du 8^{ème} colloque international PAC (Phonology of Contemporary English) à l’Université de Toulouse II-Le Mirail (29 février-2 mars).

Communications orales dans colloques nationaux avec CL

2022

- **DABOUIS, Q.** (2022) “Failles (in)visibles ? Morphologie opaque et phonologie en anglais”, communication à l’atelier de l’ALAES et de l’ALOES du 61^{ème} congrès de la SAES à l’Université Clermont Auvergne (2-4 juin).

2018

- **DABOUIS, Q.** (2018) “Why r[e]volution? A study of the pronunciation of vowels with secondary stress”, communication présentée à l’atelier de l’ALAES et de l’ALOES du congrès de la SAES à l’Université Paris Nanterre (7-9 juin).

Communications orales dans journées d’étude ou séminaires

2024

- **DABOUIS, Q.** & **GABILLET, M.** (2024), “Voyelles finales, sous-systèmes lexicaux et accentuation des noms en anglais”, communication réalisée à la journée d’étude “La phonologie en système(s)” à l’Université de Tours, (13 juin).
- **ARNDT-LAPPE, S., SEILER, A., DABOUIS, Q.** & **GABILLET, M.** (2024) “Predicting stress in English verbs: analogy, morphology, and the level of abstraction of phonological patterns, présentation invitée au séminaire du Phonology Reading Group, University of Bolzano / University of Tromsø (21 février).

2023

- **ARNDT-LAPPE, S., DABOUIS, Q., GABILLET, M.** & **SEILER, A.** (2023) “English word stress and the Mental Lexicon – aims and first results from the ERSaF project”, présentation à l’atelier du groupe de recherche Patterns in Spoken Communication, Université de Trèves (23 novembre).
- **ARNDT-LAPPE, S., DABOUIS, Q., GABILLET, M.** & **SEILER, A.** (2023) “The ERSaF project: Aims and first results”, communication au séminaire de recherche du LRL à l’Université Clermont Auvergne (21 novembre).
- **ARNDT-LAPPE, S., SEILER, A., DABOUIS, Q.** & **GABILLET, M.** (2023) “Analogy in morphophonology : the case of English verb stress”, présentation invitée au Linguistisches Kolloquium de l’Université de Marburg (17 mai).
- **DABOUIS, Q.** (2023) “Trois synthèses pour l’épreuve de phonologie”, communication réalisée lors de la journée “Linguistique Hors Programme” organisée à l’Université de Rouen Normandie (17 février).

2022

- **DABOUIS, Q.** (2022) “Dictionnaires vs. Données orales : Le cas de la gémation de /r/ en anglais”, communication réalisée dans le cadre de la semaine “Data SHS” organisée à la MSH de l’Université Clermont Auvergne (8 décembre).
- **DABOUIS, Q.** & **MARTIN, M.** (2022) “La classification des suffixes : de Guierre à aujourd’hui”, communication réalisée lors du workshop « 40 ans de Guierre », en ligne (31 mars).

2021

- **DABOUIS, Q.** & **LUC, C.** (2021) “La gémation de /r/ en anglais britannique et américain”, communication au séminaire de recherche du LRL à l’Université Clermont Auvergne (15 octobre).

2020

- **DABOUIS, Q.** (2020) “Catégorie syntaxique et accentuation en anglais : le cas des dissyllabes verbe-nom”, communication au séminaire de recherche du LRL à l’Université Clermont Auvergne (17 avril).

2019

- **DABOUIS, Q.** (2019) "Etude graphophonologique des voyelles sous accent secondaire en anglais", communication au séminaire de recherche du LRL à l’Université Clermont Auvergne (21 février).

2018

- **DABOUIS, Q.** (2018) “La valeur des voyelles avec accent secondaire en anglais”, communication lors du Séminaire Orléanais de Linguistique à l’Université d’Orléans (19 juin).

2017

- **DABOUIS, Q.** & **FOURNIER, J.-M.** (2017) “Accentuation, structure syllabique et identité morphologique : le cas des verbes anglais”, communication réalisée lors des journées d’études du Laboratoire Ligérien de Linguistique (16 novembre).

2015

- **DABOUIS, Q.** (2015), “Structure et accent secondaire des mots dérivés en anglais”, communication réalisée lors des journées d’études internationales des jeunes chercheurs à l’Université de Tours (26-27 mars).

2014

- **ABASQ, V., DABOUIS, Q., FOURNIER, J.-M., GIRARD, I., LAMPITELLI, N., MARTIN, M.** (2014), “Combinatoire suffixale en anglais”, communication réalisée lors des journées d’études du Laboratoire Ligérien de Linguistique à l’Université de Tours (11 juin).

2012

- **ABASQ, V., DABOUIS, Q., FOURNIER, J.-M., GIRARD, I., VANHOUTTE, S.** (2012), “Présentation du poster de MFM 2012”, intervention lors de la journée d’étude “Corpus et phonologie de l’anglais : du lexique au discours” à l’Université de Tours (6 juin).

Communications comme invité

2017

- **DABOUIS, Q.** (2017) “English Weak Preservation: Frequency and Prefixation”, communication en tant que conférencier invité pour le English Linguistic Circle à l’université de Trier, Allemagne (7 juin).

Projets financés

Co-responsable scientifique du projet ANR-DFG “ERSaF” (“English Root Stress across Frameworks” – ANR-21-FRAL-0001-01), avec Sabine Arndt-Lappe (Universität Trier), de septembre 2022 à août 2025.

- Page web du projet : https://lrl.uca.fr/projet_du_lab/projet-ersaf/
- Membres associés du projet : Ingo Plag, Jean-Michel Fournier, Birgit Alber, Pierre Fournier, Tammy Ganster, Patrycja Strycharczuk
- Encadrement de deux thèses de doctorat (Marie Gabillet et Aaron Seiler).
- Recherches sur l’accentuation des noms et des verbes sans morphologie productive en anglais sur données dictionnaires et expérimentales, à l’aide de plusieurs méthodes d’analyse quantitative (régression, modèles analogiques, Naive Discriminative Learning), en échangeant avec différents cadres théoriques.
- Rencontres avec différents chercheurs spécialistes du sujet : Ricardo Bermúdez-Otero, Claire Moore-Cantwell, Canaan Breiss, Rebecca Treiman.
- Organisation de journées d’études les 10 et 11 octobre 2024 sur le thème « Nonuniformity in morphophonology across Frameworks » à l’Universität Trier.

Sociétés savantes

Membre de l’ALOES (depuis 2014) et de la SAES (depuis 2018).

Autres activités de recherche

1. Organisation de conférences/journées d'études

- Membre du comité d'organisation des journées d'études « Nonuniformity in morphophonology across Frameworks » à l'Universität Trier (Allemagne, 10-11 octobre 2024).
- Co-organisateur de la journée d'étude « La phonologie en système(s) » à l'Université de Tours, (13 juin 2024).
- Membre du comité d'organisation du 18^{ème} colloque international du Réseau Français de Phonologie (RFP) (1^{er} – 2 juillet 2021, Clermont-Ferrand).
- Membre du comité d'organisation des premières rencontres des « héritiers » de Lionel Guierre à l'Université de Tours (14-15 juin 2018).
- Membre du comité d'organisation de la troisième rencontre du groupe de recherche international Phonological Theory Agora (PTA) à l'Université de Tours (14-15 octobre 2016).
- Membre du comité d'organisation des journées d'études internationales des jeunes chercheurs en linguistique à l'Université de Tours (26 et 27 mars 2015).
- Membre du comité d'organisation de la journée d'étude « Corpus et phonologie de l'anglais : du lexique au discours » à l'Université de Tours (6 juin 2012).
- Membre du comité d'organisation du 9^{ème} colloque international du Réseau Français de Phonologie (RFP) à l'Université de Tours (30 juin – 2 juillet 2011).

2. Activités de relecture

- **Revues :** *Morphology*, Springer, *Canadian Journal of Linguistics*, Cambridge University Press, *Journal of Linguistics*, Cambridge University Press, *Radical*, CNRS, *Glossa: a journal of general linguistics*, *English Language and Linguistics*, Cambridge University Press, *SKASE Journal of Theoretical Linguistics*, Slovak Association for the Study of English, *Cahiers du LRL*, Presses Universitaires Blaise Pascal, *Lexis: Journal in English Lexicology*, CEL, Université Jean Moulin Lyon 3.
- **Participation à des comités scientifiques :** *Lexis: Journal in English Lexicology* (depuis 2023)
- **Conférences :** Réseau Français de Phonologie (RFP), Generative Linguistics of the Old World (GLOW 46), Old World Conference in Phonology (OCP 20), Workshop “Nonuniformity in morphophonology across Frameworks”.
- **Ouvrages :** ARNDT-LAPPE, S., BRAUN, A., MOULIN, C. & WINTER-FROEMEL, E. (eds.) (2018), *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and the Role of Discourse-related Factors*. Berlin, New York: de Gruyter Mouton.

3. Edition

- Membre du comité éditorial de la revue en ligne *Radical* (<https://radical.cnrs.fr>)
- Co-direction d'un numéro de *Lexis: Journal in English Lexicology* à paraître en 2024 sur les liens entre lexicologie et phonologie, avec Christophe Coupé, Olivier Glain et Vincent Hugou.
- Edition numérique de la thèse de Lionel Guierre (1979, *Essai sur l'accentuation en anglais contemporain : Eléments pour une synthèse*), avec Marjolaine Martin (mars 2022).
- Co-éditeur des *Proceedings of the LLL Young Researchers Conference 1*, avec Augustin Ndione et Camille Ternisien, 2016. URL : <https://hal.archives-ouvertes.fr/hal-01379933v1>

ENCADREMENT D'ETUDIANTS

Co-encadrement de thèses

- MARIE GABILLET (avec Damien Chabanal, Université Clermont Auvergne & Sabine Arndt-Lappe, Universität Trier) *L'accentuation des noms en anglais britannique contemporain* (depuis le 1/10/2022), dans le cadre du projet ERSaF.
- AARON SEILER (avec Sabine Arndt-Lappe, Universität Trier) *Stress placement in three-syllabic nonce verbs: Evidence from speech data* (depuis le 31/08/2022), dans le cadre du projet ERSaF.

Encadrement de mémoires de master

- JULIEN PENARD. *Étude Phonologique sur l'accentuation des paires nom-verbes dissyllabiques en anglais*. Mémoire de M2 LLCER commencé en septembre 2023.
- YANN BIANCONI. *Sculpting Soundscapes: Unravelling the Enactive Power of Phonaesthemes through the Poetic Thread of a Corporeal Tapestry*. Bilan de recherche de M1 LLCER rendu en juin 2024.
- ELISE HIBON. *Consonant clusters and the notion of strength when it comes to their resistance to phonetic erosion*. Bilan de recherche de M1 LLCER rendu en juin 2024.
- ANAÏS RENE (Université de Tours, co-direction avec Marjolaine Martin). *Le flapping en anglais américain : Étude de corpus des environnements /nt/, /lt/ et /rt/*. Mémoire de M2 Linguistique soutenu en juin 2024.
- MÉRYL SAMMARCELLI. *The Super Power Of Prefixes: A Semantic Study of The Augmentative Prefixes in the English Language*. Bilan de recherche de M1 LLCER rendu en juin 2024.
- RIHAB ZOUAOU. *From Hollywood to High Streets: Is American English Reshaping British English?* Bilan de recherche de M1 LLCER rendu en mai 2024.
- YOUSSEF FADEL. *Stress placement in English disyllabic verbs and nouns*. Mémoire de M2 LLCER soutenu en juin 2023.
- CORENTIN LUC
 - *Language and video games: the influence of accents on Overwatch players*. Mémoire de M2 LLCER soutenu en septembre 2022.
 - *Language and video games: the influence of accents on video game players*. Bilan de recherche de M1 LLCER rendu en juin 2021.

Suivi de stages de master LLCER parcours Médias et médiation culturelle

- NONI CHARLENE. *Définir des stratégies de communication dans le milieu associatif*. Rapport de stage de M1 LLCER rendu en mai 2024.
- ARORA ARJUN. *Cultural mediation & communication*. Rapport de stage de M2 LLCER rendu en septembre 2022.
- BESSON ESTELLE. *L'importance du digital dans la stratégie de communication des entreprises*. Rapport de stage de M1 LLCER rendu en mai 2021.
- CHARKAOUI NORA. *Rapport de stage*. Rapport de stage de M1 LLCER rendu en mai 2020.

PARTICIPATION À DES JURYS DE SOUTENANCE

- BEN HANNACHI ASMA. *A Multimodal Analysis of EFL Learners' Pragmatic Competence through Virtual Exchange* (dir. Ciara Wigham). Mémoire de M2 LiDiFLES soutenu en juin 2021.
- JIN YISHAN. *L'acquisition des consonnes occlusives orales du français par des apprenants sinophones de groupes dialectaux wu et non-wu* (dir. Damien Chabanal). Mémoire de M2 LiDiFLES soutenu en octobre 2020.

COMITES DE SUIVI DE THÈSE

- MOUQUET MARIE. *How do L1 French learners perceive, pronounce and process L2 English variants? A focus on English syncope*. (dir. Herment Sophie (Université Aix-Marseille & Mairano Paolo (Université de Lille)).

AUTRES ACTIVITÉS

Participation à plusieurs associations dans les domaines de l'écologie et de l'économie sociale et solidaire :

- **Alternatiba 63** : Mobilisations citoyennes sur le climat, organisation d'événements, participation à l'édition de trois éditions d'un « Guide des alternatives écologiques et solidaires ».
- **ADML 63** : Mise en place du dispositif « Soli'doume », reliant la monnaie locale du Puy-de-Dôme aux questions d'accès à l'alimentation de qualité pour tous via un dispositif de redistribution inspiré du projet de Sécurité Sociale Alimentaire et un accompagnement aux changements de pratiques alimentaires.
- **Fresque du climat** : Animation d'ateliers collaboratifs sur le dérèglement climatique. Une vingtaine d'ateliers animés avec des publics variés (scolaires, étudiants, cadres de l'université, enseignants-chercheurs, grand public, famille...).
- **Eau'dyssée** : Animation d'ateliers de la « Fresque de l'eau » pour sensibiliser au fonctionnement du cycle de l'eau et son évolution sous l'effet des activités humaines.

Batterie depuis 2005, chant depuis 2019. Une centaine de concerts joués en France et en Europe avec plusieurs groupes. 3 albums et 7 EPs enregistrés. Organisation de concerts. Musique Assistée par Ordinateur. Composition.

SEJOURS EN PAYS ANGLOPHONES

- 2012 – 2013** Séjour en PVT (Permis Vacances Travail) de 8 mois en Australie (dont 3 semaines en Nouvelle-Zélande).
- 2010 – 2011** Séjour de 3 semaines aux États-Unis et au Canada.
- 2009 – 2010** Séjour Erasmus de 4 mois à Coventry (Royaume-Uni).

LANGUES

Maternelle : français

Autres : anglais bilingue
 espagnol notions
 allemand notions