



**HAL**  
open science

# Prosody in English pronunciation: embodiment and metacognition

Dan Frost

► **To cite this version:**

Dan Frost. Prosody in English pronunciation: embodiment and metacognition. Humanities and Social Sciences. Université Lumière Lyon 2, 2023. tel-04734999

**HAL Id: tel-04734999**

**<https://hal.science/tel-04734999v1>**

Submitted on 14 Oct 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.



# Rapport de synthèse en vue d'obtenir l'habilitation à diriger des recherches

Spécialité : Linguistique anglaise et sciences du langage

Université Lumière Lyon 2

École Doctorale Lettres, langues, linguistique

Présentée par

Dan FROST

Université Grenoble Alpes

## Prosody in English pronunciation: embodiment and metacognition

Garante : Pr. Heather Hilton, Université Lyon 2

Jury :

Pr. Olivier Glain, Université Jean Monnet Saint Etienne (président du jury)

Pr. Sophie Herment, Aix-Marseille Université

Pr. Jean-Rémi Lapaire, Université Bordeaux Montaigne

Pr. John Levis, Iowa State University

Pr. Shona Whyte, Université Côte d'Azur

*“How often we have had the experience of hearing some foreigner speak English with perfectly intelligible consonants and vowels and with standard grammatical forms; and yet we have had the greatest of difficulty in understanding because the intonational patterns were entirely unnatural and strange to us. Moreover, we may completely misinterpret a person speaking English. Proper intonation contributes a high percentage to the total intelligibility of speech.”*

*Eugene Nida (1957: 117–118)*

*“The parts of our utterance that we stress most noticeably are the ones about which we want to be most forceful, and the kinetics of that force is felt and seen in a number of ways. A writer underlining the important parts of a message does it energetically. A speaker bobs his head and swings his arms in time with his stresses. With this pugilistic obligato to the linguistic tune, it is hardly surprising that the tune became identified with the exercise that went with it”*

*Dwight L. Bolinger (1958: 127)*

*Thank you, Heather - for all your help, encouragement and honesty.*

*And thanks to the members of the jury – it is an honour and a privilege.*

*I would also like to thank my colleagues at Grenoble Alpes University, in the Lidilem, and in the CNU who allowed me to take the time off my day job to devote myself fully to this undertaking.*

*A huge thank you to all of my teachers – past and present – teachers rarely get the thanks they deserve.*

*Thank you to the colleagues with whom I have collaborated on research and teaching projects, and from whom I have learned so much. You know who you are.*

*Thanks, well done and good luck to teachers I have helped to train over the years. I have learnt so much from you all. And remember the Arab proverb “learn a language and avoid a war”.*

*I will always be more grateful than my words here can express to my friends and colleagues in the APLIUT – we have shared so much, and you have always been there for me.*

*Most of all, thank you to my students, past and present. You are why we are here.*

*None of this would have been possible without the love, help and support of my parents, children, family and friends. My love for languages and teaching started with my parents – the best teachers I have ever had. And this was never just about work. Thanks guys.*

*Et enfin, merci à toi mon ange, pour tout.*

# Contents

<b>Contents</b>	<b>4</b>
<b>Index of figures</b>	<b>8</b>
<b>Index of tables</b>	<b>10</b>
<b>Summary</b>	<b>12</b>
<b>Résumé (en français)</b>	<b>13</b>
<b>Preface</b>	<b>14</b>
<b>General introduction</b>	<b>17</b>
<b>PART ONE Prosody in language and linguistics</b>	<b>21</b>
<b>Introduction</b>	<b>21</b>
<b>CHAPTER ONE Prosody: concepts and definitions</b>	<b>23</b>
1.1. Prosody: definitions and history	23
1.1.1. Prosodic features – physical correlates and perceptual cues	25
1.1.2. Stress	27
1.1.3. Rhythm	30
1.1.4. Isochrony	31
1.1.5. Intonation	36
1.1.6. The functions of prosody	38
<b>CHAPTER TWO First language prosody</b>	<b>41</b>
2.1. The physiological components of prosody	41
2.1.1. The vocal apparatus & prosody	42
2.1.2. Iconicity	43
2.2. Prosody and cognitive neuroscience	45
2.2.1. Visual and motor activity in the brain	46
2.2.2. Prosody and the brain	47
2.3. Prosody and age	50
2.3.1. Prosody in the womb and early infancy	50
2.3.2. Prosody, accent and aging	55
2.3.3. Prosody in bilingual infants	55
2.4. Prosody & sex	56
2.5. Prosody, accent and identity	57
2.6. Prosody and the origin of human language	58
2.6.1. The musical protolanguage and gestural protolanguage hypotheses	59

2.6.2.	Bodily rhythms and the rhythms of language _____	61
2.6.3.	The prosody of Proto-Indo-European _____	63
<b>CHAPTER THREE LX prosody _____</b>		<b>64</b>
3.1.	Social factors in LX learning _____	64
3.2.	Individual differences _____	64
3.2.1.	Age _____	65
3.2.2.	Learner profiles and cognitive styles _____	67
3.2.3.	Motivation _____	68
3.3.	Foreign-accented speech and listener judgements _____	71
3.4.	Comparing and contrasting languages: transfer and associated issues _____	74
3.4.1.	Transfer in production _____	74
3.4.2.	Transfer in perception _____	75
3.4.3.	Contrastive analysis and markedness _____	77
3.4.4.	Optimality Theory _____	78
3.4.5.	Functional Load _____	79
3.5.	From speech perception and production to memorisation & language acquisition _____	80
3.5.1.	Speech perception: motor theories _____	80
3.5.2.	Speech production: the role of feedback _____	81
3.5.3.	Memory, multimodality, and the role of haptic information _____	83
3.5.4.	Models in LX acquisition research _____	87
3.5.5.	Towards a unified model _____	88
<b>CHAPTER FOUR A contrastive phonological analysis of French and English _____</b>		<b>91</b>
4.1.	Stress and rhythm _____	92
4.1.1.	Stress at the word level in English _____	92
4.1.2.	Degrees of stress, and reduction in English _____	93
4.1.3.	Stress and rhythm in French _____	95
4.1.4.	The relative weight of cues to stress in English and French _____	97
4.1.5.	Focus, tonicity and intonation in English and French _____	101
4.2.	Vowels _____	105
4.2.1.	French vowels _____	106
4.2.2.	English vowels _____	107
4.3.	Consonants, approximants and allophones _____	109
<b>Conclusion (part one) _____</b>		<b>112</b>
<b>PART TWO Prosody in LX teaching and research _____</b>		<b>113</b>
<b>Introduction _____</b>		<b>113</b>

<b>CHAPTER FIVE English and English phonology in the French educational context</b>	<b>115</b>
5.1. World Englishes	116
5.2. The status of English as an LX in France	120
5.3. Language education in France: a complex picture	121
5.3.1. Primary education	123
5.3.2. Secondary education	125
5.3.3. The tertiary sector	127
5.3.4. The private sector	129
5.4. Teacher training and pronunciation: facts and figures	129
5.5. French as a foreign language ( <i>Français langue étrangère</i> )	131
5.6. Technology mediated pronunciation learning	134
<b>CHAPTER SIX Prosody in production, perception and comprehension: LX research</b>	<b>139</b>
6.1. Accent, intelligibility and comprehensibility	140
6.1.1. Defining the constructs	141
6.1.2. Another approach to intelligibility and comprehensibility: clinical communities	142
6.1.3. Accentedness in LX research	144
6.1.4. Comprehensibility in LX research	147
6.1.5. Intelligibility in LX research	148
6.2. Teaching prosody and learning outcomes	150
6.2.1. LX English	152
6.2.2. Other languages	156
6.2.3. The French context	156
6.3. Deconstructing intelligibility: towards a model	158
<b>CHAPTER SEVEN Prosody in LX learning, teaching and assessment</b>	<b>162</b>
7.1. Norms, varieties and models	162
7.1.1. The “native speaker debate”	163
7.1.2. Aims for LX English pronunciation	164
7.1.3. Accent addition or accent exposure (pronunciation for reception)	165
7.1.4. Listener training	166
7.2. Prosody & assessment	167
7.3. Towards an embodied and metacognitive approach centered on prosody	173
7.3.1. Examples of embodied approaches to language learning	173
7.3.2. Music, prosody and language learning	175
7.3.3. Metacognition and learning	178
7.4. An embodied and metacognitive approach centered on prosody: what and how	179
7.4.1. Towards a prosody-centered syllabus	180
7.4.2. Activities: principles	182

7.4.3. Activities: examples _____	183
7.5. CAPT: embodied cognition and computer-assisted pronunciation instruction _____	185
7.6. The ethical dimension of pronunciation instruction _____	187
7.7. Research perspectives & projects _____	188
<b>Conclusion (part two) _____</b>	<b>192</b>
<b>General conclusion _____</b>	<b>194</b>
Motivation _____	194
Context and content _____	194
Projects past and present _____	194
Learning and teaching _____	195
From embodied cognition, to embodied metacognition _____	195
Teacher training _____	196
Engagement, entrainment, and <i>engagement</i> _____	196
<b>Bibliography _____</b>	<b>198</b>
<b>Appendix: prosody-focused pronunciation instruction studies included in meta-study (6.2) _</b>	<b>255</b>



## Index of figures

Figure 1. The place of prosody in phonology	22
Figure 1.1. Steele’s annotation of “accents” (Steele, 1779: 7)	24
Figure 1.1.2a. Syllable structure, adapted from (Aslam & Aadil, 2011: 61)	28
Figure 1.1.2b. Altmann & Vogel’s Typology of stress parameters (Altmann, 2006: 31)	30
Figure 1.1.4. Levels of Temporal and Rhythmic Structuring (Kotz et al. 2018: 898)	35
Figure 2.2.2. Schematic models of shared brain network for rhythm perception and production in speech and music (Fuji & Wan, 2014: 777)	49
Figure 2.3.1a. The timeline of infant speech acquisition (from Kuhl (2004), with perceptual and motor studies on rhythmic acquisition added (Goswami, 2022: 3))	52
Figure 2.3.1b. A model of speech processing and early language acquisition (Christophe et al., 2008: 62)	52
Figure 2.6. A schematic representation of organism-external and -internal factors related to the faculty of language (Hauser et al., 2002: 1570)	59
Figure 2.6.2. The role of rhythmic cognition in speech emergence from an evolutionary perspective. (Polyanskya et al., 2019: 8)	62
Fig 3.2.3. Diagrammatic representation of the difference between efficacy expectations and outcome expectations (Bandura, 1977b: 193)	70
Figure 3.5.1. The PACT model of speech perception (Schwartz et al., 2012)	81
Figure 3.5.2a. Levelt’s model of speech production (Levelt, 1989: 9)	82
Figure 3.5.2b. Levelt’s model of speech production (Levelt, 1999: 3)	83
Figure 3.5.3a. The structure of the memory system (Atkinson & Shiffrin, 1968: 93)	84
Figure 3.5.3b. Verbal and nonverbal symbolic systems of Dual Coding Theory (Paivio, 1986: 38)	85
Figure 3.5.3c. The three-component model of working memory proposed by Baddeley and Hitch (1974) (Baddeley, 2003: 191)	86
Figure 3.5.3d. Baddeley’s current model of memory illustrating the flow of information from perception to the episodic buffer (Baddely, 2021: 870)	86
Figure 4.1.2. The four degrees of stress in British (RP) and American (GA)	94
Figure 4.1.4. Rhythm-melody (RM) space for English and French speech and music (Patel, 2006: 3041)	99
Figure 4.1.5a. The forms of the ten canonical intonation patterns in French (Delattre, 1965:4)	102
Figure 4.1.5b. Comparison between two L1 English users, two C level learners and one B level learner (Herment et al., 2014: 4)	103
Figure 4.1.5c. Examples of English and French intonation curves (Delattre, 1963: 196)	104
Figure 4.1.5d. Examples of how an L1 English user and an L1 French user would pronounce the same sentence (adapted from Delattre, 1963: 194)	104
Figure 4.2.1. French oral vowels (Fougeron & Smith, 1993: 73)	106
Figure 4.2.2a. Received Pronunciation: short and long monophthongs, or “pure vowels” (Roach, 2004: 242)	107
Figure 4.2.2b. Received Pronunciation: closing diphthongs and centring diphthongs (Roach, 2004: 242)	107

Figure 4.2.2c. Tense versus lax vowels in North American English (Celce-Murcia et al., 1996: 96)	108
Figure 5. An outline of a syncretic model of phonological acquisition grounded in phenomenological phonology (PP) (Leather, 2002: 62)	116
Figure 5.1. Kachru's (1985) three-circle model of world English (MacArthur, 1987: 11)	117
Figure 5.2. Relative emphasis teachers put on pronunciation, ESLC (European Commission, 2012b: 204)	120
Figure 5.4a. Received training in teaching the target language as a foreign language, ESLC (European Commission, 2012b: 196)	130
Figure 5.4b. Perceived quality of training to teach pronunciation. (Frost & Henderson, 2013: 102)	131
Figure 5.7. Salient features of different categories of CAPT systems for English. (Agarwal & Chakraborty, 2019: 3740)	135
Figure 6.4. An embodied cognition model of speech	160
Figure 7.1.2. Issues for negotiated LX English pronunciation learning goals (production)	165
Figure 7.3.3. Metacompetence-oriented model of phonological acquisition (Wrembel, 2007: 191)	179

## Index of tables

Table 4.1.3. A comparison of the position of stress on French and English words (adapted from Delattre 1965: 29) _____	95
Table 4.1.4a. Rhythm values for 14 languages using the Pairwise Variability Index (PVI) (adapted from Grabe & Low, 2002) _____	98
Table 4.1.5a. The ten canonical intonation patterns in French by function (Delattre, 1965:3) _____	102
Table 4.2.2d. /r/ coloured vowels in North American English (Celce-Mucia et al., 1996:105) _____	109
Table 5.1. EIL (English as an International language) and NS (native speaker) pronunciation targets (Jenkins, 2002: 99) _____	119
Table 6.3.1. The effects of different types of instruction on comprehensibility, accentedness and fluency over a 12-week ESL course. (Derwing et al., 1998: 403) _____	154
Table 7.2b. The prosody descriptors (Frost & O'Donnell, 2018) _____	170
Table 7.2c. Common European Framework of Reference for Languages - <i>Companion Volume</i> "phonological control descriptors" (Council of Europe, 2001: 28) _____	172



## Summary

Pronunciation teaching has waxed and waned in importance over the decades, as have attitudes to what and how we should teach it. Current thinking is very influenced by the constructs of intelligibility and comprehensibility, but the actual features on which we need to concentrate in any given learning situation will depend on the target language and the language(s) of the learners in question. My starting point has always been the following questions: why do many French speakers (and others) have difficulties understanding and speaking English, and what can we do about it? In this report, I explore the reception and the production of language, I outline the research which has informed my research and teaching over the last 25 years and why I believe that prosody is key to addressing these questions, and I will present my own work where it is relevant to the questions being discussed.

Pronunciation is both a physical and a cognitive skill, and so I adopt a framework of embodied cognition, situating the mind in the body, and placing both mind and body together in the wider environment. In the first part of this report, I address the questions more from the standpoint of a linguist: I define some key concepts to help us to understand prosody better. Then, I explore the evolution of human language, the development of our first language, and the role prosody plays in these processes, before dealing with some more general aspects of using and learning languages, both from an individual and a societal perspective. Finally, I examine some of the reasons why it is important and difficult to focus on prosody, especially for many French learners. The second part of this report deals with these questions through the lens of learning and teaching languages. I explain some of the ways in which my research and the research of many others has shown that focussing on prosody can be beneficial, and finally, I present some ideas on how cognitive and meta-cognitive techniques, integrated with an embodied approach to teaching can help achieve better learning outcomes.

**Key words:** cognitive; embodied cognition, embodied teaching; meta-cognitive strategies; pronunciation; prosody.

## Résumé (en français)

### **Pourquoi et comment mettre l'accent sur la prosodie dans l'enseignement de la prononciation**

*L'enseignement de la prononciation a connu des hauts et des bas au fil des décennies, tout comme les attitudes à l'égard de ce qu'il faut enseigner et de la manière dont il faut l'enseigner. La pensée actuelle est très influencée par les concepts d'intelligibilité et de compréhensibilité, mais les éléments d'une langue sur lesquels nous devons nous concentrer dans une situation d'apprentissage donnée dépendent de la langue cible et de la (des) langue(s) des apprenants en question. Mon point de départ a toujours été les questions suivantes : pourquoi de nombreux francophones (et autres) ont-ils des difficultés à comprendre et à parler l'anglais, et que pouvons-nous faire à ce sujet ? Dans ce rapport, j'explore la réception et la production du langage, et je décris les recherches qui ont alimenté mes recherches et mon enseignement au cours des 25 dernières années, et les raisons pour lesquelles je pense que la prosodie est essentielle pour répondre à ces questions, et je présenterai mes propres travaux lorsqu'ils sont pertinents pour les questions discutées.*

*La prononciation est une compétence à la fois physique et cognitive, et j'adopte donc un cadre de cognition incarnée, situant l'esprit dans le corps, et plaçant l'esprit et le corps ensemble dans un environnement plus large. Dans la première partie de ce rapport, j'aborde les questions plutôt du point de vue d'un linguiste : je définis quelques concepts clés pour nous aider à mieux comprendre la prosodie. Ensuite, j'explore l'évolution du langage humain, le développement de notre première langue et le rôle que joue la prosodie dans ces processus, avant d'aborder certains aspects plus généraux de l'utilisation et de l'apprentissage des langues, tant d'un point de vue individuel que sociétal. Enfin, j'examine certaines des raisons pour lesquelles il est important et difficile de se concentrer sur la prosodie, en particulier pour de nombreux apprenants de français. La deuxième partie de ce rapport aborde ces questions sous l'angle de l'apprentissage et de l'enseignement des langues. J'explique comment mes recherches et celles de beaucoup d'autres ont montré que se concentrer sur la prosodie peut être bénéfique, et enfin, je présente quelques idées sur la façon dont les techniques cognitives et métacognitives, intégrées à une approche incarnée de l'enseignement, peuvent aider à obtenir de meilleurs résultats d'apprentissage.*

**Mots clés :** *\_approche corporelle ; cognition incarnée ; stratégies métacognitives ; prononciation ; prosodie.*

## Preface

This report, the *note de synthèse en vue de l'obtention de l'habilitation à diriger des recherches*, is of course a requirement in France in order to supervise PhDs independently, and to aspire to the grade of full professor, but it is also a chance to take a step back and examine one's research, to take stock. Perhaps it is hindsight and a human need to find sense in one's past decisions, but the more I look at the path I have taken, the different contexts in which I have carried out my teaching and my research over the past 30 years, the more clearly I see the reasons behind the choices I have made. The things that I have learned from the people I have worked with and the places I have worked, both as a teacher and as a researcher, seem to be waypoints on a clearly marked path which has led me to the inevitable subject of this project: prosody.

I was born to Welsh parents, though I was brought up in England, my father was a music teacher, before becoming a headmaster and later an inspector, and my mother was an English teacher. During our childhood, we moved around England several times, and to aid integration in the schools and communities we found ourselves in, we all modified our accents on more than one occasion. We also spent our holidays on campsites in France, and I have happy memories of playing with other families, even though we spoke different languages. Perhaps it is unsurprising then, that my sister and I both went on to study languages and linguistics – my sister is now head of modern languages in a secondary school in the UK.

I discovered Phonetics at the University with my tutor, John Local, and phonology with John Kelly, and was fortunate enough to follow phonetics classes in Strasbourg on an Erasmus year with François Wioland. After completing a Cambridge TEFL<sup>1</sup> course in 1993, I taught English for two years in Thailand, and for a year in Sweden. Perhaps, had I stayed in Thailand, I would have focused more on segmental phonology, or in Sweden, more on lexical acquisition and fluency, but my experience over the next few years in France, the people I worked with and the research I read and conducted myself led me in another direction. I arrived in France in 1996 at what was then the University of Aix-Marseille as a teaching assistant in English phonetics (*lecteur en phonétique anglaise*), where I was fortunate enough to have colleagues such as Michel Ginésy, André Lipcey, Philippe Dominique and Michel Corsi. All of these exceptional teachers and researchers had studied English under Georges Faure, and it was his *Manuel Pratique d'Anglais Parlé* (Faure, 1948) which I used to supplement the courses in English phonetics I taught using Michel Ginésy's course books (Ginésy, 1989; 1995). These years were formative years in my career, and the influence of Faure's work could and can still be felt in the work

---

<sup>1</sup> Teaching English as a Foreign Language (now known as the CELTA course).

carried out in the University of Provence, both in the teaching and in the research: in 1971, Georges Faure wrote “It is better to approach the teaching of a foreign language from the analysis of prosodic structure than from phonematic ones” (Faure, 1971a: 7). During my first year as a teaching assistant, I completed a *Maîtrise d'anglais (mention phonétique)*, a collaboration between the English and Linguistics departments, under the supervision of Dan Hirst and Albert Di Cristo. This involved following classes mainly in Phonetics and Phonology, carrying out my first experiments in perceptual phonetics and reading research papers, mainly on prosody, as the Laboratoire Parole et Langage<sup>2</sup> has always focused on prosody since it was founded by Georges Faure in 1962.

The following year saw my focus shift from research to teaching, when I took the exams necessary to become a teacher in France: first the *CAPES externe* in 1998, then following my posting to Normandy, the *Agrégation externe* in 2002, and my choice to specialise in *la linguistique* allowed me to further explore the relationship between teaching and English phonology. Whilst teaching and studying for the *Agrégation* in Rouen, I started a *DEA*<sup>3</sup> at the Bordeaux II University, and taught at the IUT2 (*Institut universitaire technologique*) de Grenoble. This position allowed me to return to my research, and I signed up to do a PhD by distance learning at Bordeaux, supervised by Jean-Louis Duchet. For my PhD, I explored the use of digital technologies to improve the acquisition of English by concentrating on prosody (Frost, 2004; 2008), and on defending my PhD, I obtained my first associate professor post in the Savoie Mont Blanc University, where I was teaching phonetics and oral acquisition of English vocabulary to first years in the Applied Foreign Languages department. The programmes I was teaching had been pioneered by my predecessor, Heather Hilton, and were a pragmatic solution to the real problem of a surprisingly low level of English among the first-year students; in a later study, my colleague Jean O’Donnell and I found their level in oral production and oral interaction to be an average of A2 on arrival at our university (Frost and O’Donnell, 2013; 2015).

In 2015, I joined Grenoble Alpes University on a post involving mainly teaching adults, and Lifelong Learning has been a common feature of my career: I have always either taught adults full-time, or taken extra hours teaching adults alongside the teaching I have done with younger learners. Since my arrival in Grenoble, I have been able to put my experience with teaching learners of all ages and in many different contexts to use through teacher training, where I continue to focus on pronunciation in my teaching, teacher training and research. While it is true that there has been a renewed interest in pronunciation in second language acquisition (SLA) research in the last two or three decades, this

---

<sup>2</sup> <https://www.lpl-aix.fr/en/welcome-to-lpl/>

<sup>3</sup> *Diplôme d’études approfondies* – a pre-doctoral programme, or what would now be the second year of a Master’s programme.



has not necessarily carried through to teaching in practice. As I found out when I was part of a team that developed a survey (EPTiES, English Pronunciation in Europe Survey), most teachers of English have little or no training in how to teach pronunciation at all (Frost & Henderson, 2013; Henderson, et al., 2015). There are many ways to teach pronunciation, and they are all more or less successful, however not teaching pronunciation (or not learning how to teach pronunciation) is unlikely to produce favourable results. For this reason, much of my effort in the last decade has gone into teacher training, and I will describe a few of these projects at the end of this report.

In this report, I will draw on the work of many colleagues who have conducted research in a large variety of contexts, and I will present my own research and experience as a teacher and teacher trainer where relevant. Several research and teaching projects have enabled me to conduct original research and make a few modest contributions to the field, as well as to meet and exchange with some of the most accomplished researchers nationally and internationally, in applied phonetics and phonology, and in teaching and researching pronunciation. In addition, I have co-supervised, and continue to co-supervise doctoral students, which have enabled me to explore other aspects of learning languages, to pass on some of what I know and to learn more. And of course, the many Masters students I have supervised over the past fifteen years, many of whom have specialised in pronunciation in one form or another, continue to inspire and impress me. The privileged position of accompanying people on the first part of their journey to becoming teachers and researchers is something I will always be grateful for – and that is my main motivation for applying for the *Habilitation à diriger des recherches*.

It is both a disconcerting and a comforting experience to review the entire output of one's career as a researcher. There are times when I have read my earlier publications and felt a degree of shame at the naïve and hastily-formed opinions, the errors in approach and analysis, in short, at my own lack of knowledge. As the *cliché* goes, the more I learn, the more I realise I don't know, and although I can see a trajectory through my work, there is much repetition and much reiteration of similar ideas, and I know there will never be enough time in my career to explore all the subjects I wish to explore. When I read my own work however, I also see that it was mostly relevant to what I was reading and teaching at the time, as I hope my work is now and always will be.

## General introduction

The question which has been at the heart of both my teaching and my research since I arrived in France has been “why do so many French people have difficulty understanding English when it is spoken by native speakers?” In fact, the reason I have chosen to focus on pronunciation in my work is not an attempt to improve the pronunciation of French learners of English for the sake of it, but because it has become increasingly clear to me both through my research and the research I have read – in addition to my experience of teaching and teacher training – that pronunciation is at the heart of this problem.

Before addressing the question of pronunciation and the role of prosody in language and language learning, it would be useful to examine several terms often used to describe language learning and teaching. Although my career has involved many different teaching situations, the context is now almost exclusively in France. Most of my research and my teaching is therefore geared towards teaching English as a *foreign* language (TEFL), and not teaching English as a *second or other language* (TESOL). That being said, I am increasingly unsatisfied with this dichotomy and with the opposition of “foreign” and “second” reflected in the terms TEFL and TESOL. In much of the English-speaking world, many of us only have one first language (L1), and often only one “foreign” language (L2), however this is of course not the norm worldwide, or even in Europe. Many of my students here in France have gone on to live in English-speaking countries, so were we teaching them ESL and not EFL all along? Or should we have been? After all, there is more and more geographical mobility, and virtual mobility in this ever-shrinking world. We increasingly teach people who, for whatever reasons, are in France, but were not born here. Covid-19 impacted years aside, there are more and more international students in France – 10% according to the French Ministry of Education (Javet, 2021) and for most of these students, English is not their first language – and then of course there are all of the students from North Africa, and West and Central Africa, who spoke other languages before being exposed to French. In my work in lifelong learning, many of my students were born outside France and arrived here for work reasons, or because their parents or partners chose to live here. In fact, although English is the most widely spoken foreign language across the European Union according to the last major European survey on language use (European Commission, 2012: 24), English is often the third, or even fourth or fifth language of learners. As for the terms “native speaker” and “non-native speaker”, these are still widely used in research, and I still use them myself occasionally for convenience, though they are also problematic for several reasons. The word “native” may be negatively connotated, and may be seen to imply a degree of mastery, or at least literacy which certain language users do not have. But perhaps more importantly for teaching pronunciation, “nativism” (Levis 2008; 2020) - teaching pronunciation towards native a native-speaker model, and ultimately towards native-speaker like proficiency - is still

considered a useful and achievable goal by many teachers and learners, and we will discuss this issue and its implications in chapter seven. Finally, the term “speaker” is also problematic, not least because much of the contact people have with a language, especially English, is not actually speaking. For these reasons (unless reporting studies where these terms are used by authors) I will use the term “user” rather than “speaker”, and “LX”, rather than “L2”, where “X” refers to “any foreign language acquired after the age at which the first language(s) was acquired, that is after the age of 3 years, to any level of proficiency” (Dewaele, 2017: 238).

Interest in researching and teaching pronunciation has waxed and waned over the years, waning when the communicative approaches focused on fluency at the beginning and end of the twentieth century, waxing when there was a pushback in the research community, or when analogue or digital technological advances allowed new possibilities in the laboratory or in the classroom. But there has undoubtedly been a resurgence of interest in pronunciation in the last two decades (Pennington, 2019), as we can see by the creation of the annual conference in North America, *Pronunciation in Second Language Learning and Teaching (PSLLT)* in 2009, the *Journal of Second Language Pronunciation (JLSP)* in 2015, the creation of the annual *Accents* conference in Łódź, Poland, in 2009, and *English Pronunciation: Issues & Practices (EPIP)* here in France, in 2008. I have been part of these research communities since I completed my PhD in 2008, and co-organized EPIP7 in Grenoble in 2022. Rather fortuitously then, the recent rise in interest in research in teaching pronunciation and the research-based decisions regarding the way pronunciation is increasingly taught, coincide with the 25 years of my career here in France. The themes in research on teaching pronunciation have also shifted during this time, compared to attitudes in the twentieth century and before. Pennington (2021) lists the following trends which have marked pronunciation teaching over the past three decades:

1. greater acceptance of the mutability of pronunciation, and not only in childhood;
2. greater understanding and acceptance of the relevance of social and psychological factors to pronunciation;
3. multilingual orientations to pronunciation;
4. intelligibility and communicative effectiveness as goals rather than correctness or accuracy;
5. communicative and task-based methodologies that combine a focus on meaning with a focus on form;
6. pronunciation taught in specific communicative contexts;
7. high attention to advanced learners and those with employment-related needs in pronunciation;
8. and continuing development and improvement of applications of electronic resources to pronunciation. (Pennington, 20: 3-4).

In the first part of this report, I will address questions which Pennington raises concerning some of the individual and societal issues relating to learning and teaching pronunciation (points 1-3), and in the second part, we will deal with the more pedagogical trends she mentions (4-8). If we look at the past 50 years of research into teaching pronunciation, prosody has clearly been neglected in favour of segmental features. In a survey of major international journals in L2 acquisition between 1969 and

2008 (Gut, 2009), it was found that of 133 empirical studies on L2 phonology, only 17 pertained to prosody. This balance has shifted in the past decade or so (Mennen & der Leeuw, 2014), but phonemes are still the focus of many researchers, and are treated far more often than suprasegmental features in the classroom. In recent years, the subject of whether to focus on segmentals or suprasegmentals is often discussed (Wang, 2022). While I believe this in some ways to be a false dichotomy, it is still a relevant question, in that at any one time, choices must be made on what to teach. However, as I hope to make clear, segmental and suprasegmental features should not be treated separately. I prefer to foreground prosody, and to look at phonemes through the way prosody affects them wherever possible, as prosodic features very often have a bearing on the behaviour of phonemes. If we compare prosody to music, then we can see many similarities (Patel, 2010), especially in the way meter structures both language and music (Palmer and Kelly, 1992; Fedorenko et al., 2009).

In most English departments in France, phonemes are studied before prosody, and given much more prominence in the programmes. As I mentioned in the preface, my first job in France was as a *lecteur de phonétique anglaise*, and the programme for the first-year undergraduates studying English was learning to recognise, reproduce and transcribe the vowels and consonants of Received Pronunciation (RP) and General American (GA). Primary word stress was dealt with a little, but sentence stress, intonation, secondary stress, etc. were hardly mentioned until the second year. My research and teaching have increasingly led me to realise that this is putting the cart before the horses. Of all the aspects of speech, I believe prosody to be the most physical, the most iconic, and the most fundamental of all, as we shall see in the first chapter of this report. I am not alone in placing an importance on prosody in my research: in the introduction to his collection of his papers in honour of George Faure, Di Cristo (2004) comments on the recent growth in attention to prosody in research across several domains of linguistics and applied linguistics:

*Cette propagation soudaine de l'intérêt pour la chose prosodique, pourtant si souvent négligée naguère (au point qu'on a pu la qualifier de « cendrillon de la communication », cf. Bolinger, 1986 ; Fónagy, 1989), pourrait passer pour une simple épidémie de prosodomania (ce qui est partiellement vrai), si elle n'était en réalité porteuse d'un esprit d'ouverture salutaire qui favorise un décloisonnement des disciplines et qui attise ainsi la volonté d'appréhender les phénomènes linguistiques dans l'intégralité de leur espace cognitif.*

The sudden growth in the interest for all things prosodic, hitherto so often neglected (to the point where it could have been referred to as the “Cinderella of communication”), cf. Bolinger, 1986 ; Fónagy, 1989), may be considered a simple epidemic of prosodomania (which is not entirely untrue), if it were not in reality responsible for an opening of the minds favouring a breaking down of barriers between research communities, thereby rekindling the will to understand linguistic phenomena in the entirety of their cognitive space.<sup>4</sup> (Di Cristo, 2004: 68).

---

<sup>4</sup> I have translated all of the quotations in the report into English where necessary.

As I have discovered through researching prosody over the last twenty-five years, and as we shall see in this report, prosody is highly iconic, and is strongly related to body movements, expressions and gestures. It contains many features which are common across languages, and it is one of the oldest features of human language. It is the first thing we acquire in our mother tongue, and it is one of the last features of our mother tongue we lose. It is essential for speech processing, and conveys morpho-syntactic, semantic, pragmatic, lexical information. And yet prosody is often neglected or poorly taught in the language classroom. I hope in this report to demonstrate the importance of foregrounding prosody when teaching language, when training teachers and researchers, and when conducting research in the field of language acquisition, learning and teaching. More than two decades of research, teaching, and training teachers – particularly in English pronunciation – has increasingly led me to explore the role of the body and its relationship with the mind and the environment, an approach which has variously been referred to as embodied cognition (Varela et al., 1992), and grounded cognition (Barsalou, 2008). In the first part of this report, we shall examine the nature and roles of prosody, and address some of the questions of prosody in one's mother tongue. We will also present some of the theories and models relevant to teaching and researching how prosody is learned, especially by French learners of English. The three chapters in the second part of this report will enable us to explore the role of various external factors and how they influence prosody, how prosody contributes to various issues in language production and comprehension, and finally the implications for learning, teaching, and assessing English. Where relevant, I shall mention my own research, and some of the teaching and teacher-training projects I have been involved in over the past 25 years.

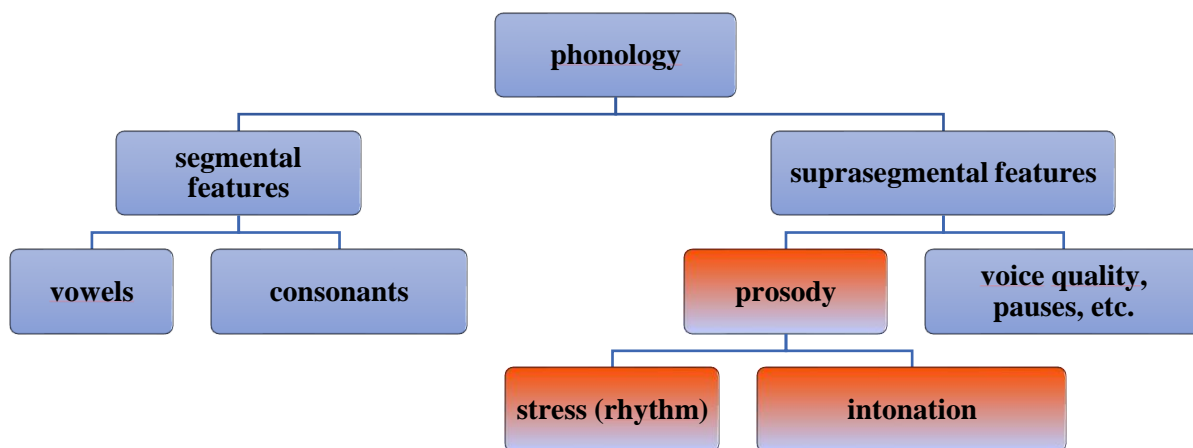
## PART ONE

### Prosody in language and linguistics

#### Introduction

It is not my aim in this part to offer an exhaustive description or definition of prosody, which would be beyond the scope of this report. As Elizabeth Couper-Kuhlen (1986:1) states: “Terminological confusion is rampant in many fields of scientific inquiry but it is perhaps unmatched in the field of prosody”. Authors do not always agree on a definition of prosody, either in terms of its nature or its functions. I will therefore, attempt to concentrate on those aspects of prosody which are most central to my work with French learners, especially stress and rhythm, and how they affect the learning of English as an L2, especially (but not only) for L1 French learners. In this part of the report, I will address prosody more generally, dealing mainly with prosody in our first language, and in the second part, more from the point of view of teaching and learning English.

While it is true that the terms “prosody” and “suprasegmental” are often used interchangeably, prosody is generally considered to be a narrower term. Suprasegmental features may include all features which transcend segments (phonemes), and may include pauses, paralinguistic features such as voice quality, etc., and prosody is limited to intonation and stress (and therefore rhythm), and tempo. If we say that at its simplest, prosody is a question of stress and intonation, we oppose these two concepts, yet they are quite clearly linked: variation in intonation is achieved by changing the speed at which the vocal cords (or vocal folds) vibrate (F0), and F0 is also one of the four acoustic correlates for realising stress, as we shall see later. As one of the other correlates is the duration of the stressed syllable, some syllables are longer than others, which gives rise to changes in tempo and rhythm. So, for the purposes of this report, I will align myself with most authors on the subject, and use the term prosody to refer to stress (and therefore rhythm and tempo, which are governed by stress) and intonation. The following diagram represents prosody and the place it occupies in phonology as I will be using it in this report.



**Figure 1. The place of prosody in phonology**

While prosody is often considered to be completely separate from segmental features (as one might assume from diagrams such as figure 1), they are intrinsically related. This is especially true in a language such as English, where stress and rhythm have a major influence over the nature of phonemes. Prosody and phonological skills are, by nature, a cognitive *and* a physical – be it on the level of production or reception – and prosody in particular is an embodied form of communication: of all aspects of oral language, prosody is the domain where form and meaning most coincide, and the rhythms of our body frame the rhythms of our language. Figure 1 represents the different domains of study within the field of phonology, but it does not account for the place of visual signs. While auditory signs convey much of the information in the process of speech communication, they are accompanied by visual signals, and these visual signals play a crucial role in phonological features. The first part of this report will deal with prosody in general, and with the prosody of English, especially as it affects the acquisition and use of our L1. The first chapter will address the nature and functions of prosody, which will allow us to examine how prosody affects learning languages, and finally, in the third chapter, the particular issues of learning English, especially for L1 French users.

## CHAPTER ONE

### Prosody: concepts and definitions

In this chapter, we explore prosody generally, without broaching issues pertaining to language learning. We will start with some definitions of prosody, and of its constituent parts, before briefly examining the functions of various prosodic features. This chapter is not intended to deal with such a complex subject in an exhaustive manner, but to give an overview of the main phenomena which my research actively addresses. By defining the phenomena themselves and their characteristics in English, we will better be able to discuss the problems they pose for learners, teachers and researchers in the context of English as an LX in France.

#### 1.1. Prosody: definitions and history

It is very hard to describe one prosodic feature without referring to other features, as many of them work together and influence each other. The definition of the term “prosody” in David Crystal’s *Encyclopaedia of Linguistics and Phonetics* begins:

A term used in suprasegmental phonetics and phonology to refer collectively to variations in pitch, loudness, tempo and rhythm. Sometimes it is used loosely as a synonym for ‘suprasegmental’, but in a narrower sense it refers only to the above variables, the remaining suprasegmental features being labelled paralinguistic. (Crystal, 2008: 394).

Prosody essentially refers to two things: rhythm (the beats of language, whether syllables are “stressed” or “unstressed”), and intonation. Pauses, generally considered a suprasegmental feature but not a prosodic feature, do have bearing over prosody, in that they separate speech into larger prosodic units. In this chapter, and throughout this report, I will sometimes refer to prosodic features across languages, but I will focus on the prosody of English.

The word “prosody” itself comes from the Greek word *προσωδια* referring to the way songs were sung to music, or the tones of syllables, or the “laws of metre” (Noteboom, 1997). There are references to the “melody” and “music” of English in Cooper’s *The Discovery and art of Teaching and Learning the English Tongue* (Cooper, 1687), and various prosodic features are often mentioned in the series of dictionaries and pronouncing dictionaries published in the second half of the eighteenth century.



These pioneering dictionaries were as prescriptive as they were thorough, as the title of John Walker's pronouncing dictionary (1791) explicitly states<sup>5</sup>.

However, perhaps the most authoritative work on English prosody of this period was Joshua Steele's *Prosodia Rationalis* (Steele, 1779). The five characteristics he uses to describe prosody are "accent" (i.e. tones of syllables: rising, falling, etc.), "quantity" (syllable duration), "pause", "poize" (essentially degrees of stress), and "force" (loudness of a syllable). This remarkable treatise, although written for the president of the Royal Society, is aimed at a wider audience: it is eminently pedagogical in nature, and written with much humour, and care to make such a complex area accessible. Many of Steele's annotations and "peculiar symbols" are still used today, both in scientific writings and in teaching resources, such as the annotation of the tone on stressed syllables:





**Then, instead of using round or square heads for the notes to be marked on this scale (as in the ordinary music) let us substitute *sloping or curving lines*, such as the expression may require; as , or , or , or ; which lines, when drawn on the foregoing scale, will easily shew through how many quarter tones the voice is to slide; and these I call the *accents or notes of melody*.**

Figure 1.1. Steele's annotation of "accents" (Steele, 1779: 7)

Steele draws many parallels with music in his approach to analysing prosodic features and to annotating them. He borrows terms from music, uses staves, and refers to musical instruments with descriptions and even diagrams in his attempts to explain various features of prosody. As we will see in chapter seven, musicology is still very relevant today for explaining these concepts to learners.

Much of the work on prosody in the twentieth century focused on intonation. During this period, three main schools of thought dominated the literature on intonation according to Wennerstrom (2001): (1) Halliday's (1967a, 1967b) research on British English, (2) Pierrehumbert's (1980) dissertation and co-authored work (Pierrehumbert & Hirschberg, 1990) on intonation, and (3) Brazil's (1975, 1978, 1997) discourse-based approach to (British) English intonation. The various approaches to the analysis of prosody have coalesced into two main schools: the British tradition, and the American approach. The former tends to fall into two major categories: (1) the tune analysis, or the whole tune approach, and (2) the tonetic analysis, or the nuclear approach (Chun, 2002). The American tradition is more

---

<sup>5</sup> Walker, J. (1791) *A critical pronouncing dictionary and expositor of the English language. To which are prefixed, principles of English pronunciation. Likewise rules to be observed by the natives of Scotland, Ireland, and London, for avoiding their respective peculiarities, and directions to foreigners for acquiring a knowledge of the use of this dictionary; the whole interspersed with observations, philological, critical, and grammatical.* Robinson.

influenced by structural and generative linguistics, and is more concerned with the functions of prosody (Kang et al., 2021). In the French context however, it is rhythm and stress which are of more importance than intonation, for reasons we shall see in the last chapter of this part. Following Chomsky and Halle's *Sound Patterns of English* (1968), there were a series of books written in France, mainly aimed at undergraduate students of English (Greven, 1972; Guierre, 1984; 1987; Ginésy, 1989; 1995; Lilly & Viel, 1998a; 1998b), but which had a great influence on the way English prosody was and still is, approached by researchers in France.

Although we frequently oppose segmental and suprasegmental features, prosody and other suprasegmental combine with segmental features to make up a complex and interdependent system, which, along with non-verbal cues and a host of other linguistic and non-linguistic cues, enable the co-interlocuter to reconstruct meaning. However, prosody on its own, may provide a wealth of information, as numerous studies using low-pass filters, which remove the segmental cues and leave only the intonation, rhythm and pauses, have shown (Komatsu et al., 2002). Indeed, when prosodic patterns are ambiguous, experiments with a low-pass filter have shown that gestures help listeners to identify sentences correctly (Guellai et al., 2014). The next few sections will briefly outline the constituent parts of prosody and some prosody-related issues which are relevant to the work I present in this report.

### *1.1.1. Prosodic features – physical correlates and perceptual cues*

The prosodic features which are most consistently used to analyse and describe speech are **pitch**, **length** and **loudness** (Cruttenden, 1997:2). To these features, I would add a fourth, **timbre**, as this is an important feature in English, where unstressed syllables are often reduced. However, before we describe the nature of prosodic phenomena which interest us the most, it is important to understand the difference between measurable, physical, acoustic reality, and what a listener perceives. Pitch is the varying height of the tone that a listener hears over a syllable or a number of successive syllables, length concerns how long a listener perceives phonemes and syllables to be, and loudness concerns the changes in loudness perceived by a listener. Timbre is how a listener perceives the difference between individual sounds, such as /e/ and /æ/. These four *cues* are what the listener *perceives*, but what we hear does not always correspond to a measurable acoustic reality: we may easily be misled by a sound, or by its phonetic environment. When teaching intonation, for example, learners often perceive a tone as rising, when in fact it is falling, simply because it is a high or extra-high tone. And the length of a vowel in a syllable ending in a fricative such as /s/ may seem longer to a listener than a

syllable ending in a plosive such as /t/<sup>6</sup>. So, what we refer to as pitch, length, loudness, and timbre are actually **percepts**, of corresponding physical correlates, which are objectively measurable, and which are, respectively, **fundamental frequency (F0)**, **duration**, **amplitude** (or **intensity**), and **formant structure**. F0 is the frequency at which the vocal chords vibrate, and in the human voice, it is generally situated between 120 Hz and 300 Hz. Duration is measured in milliseconds, and although it may seem simple enough to measure on a spectrogram, we have to decide where the syllable break occurs, and exactly where a vowel starts and ends. Amplitude (or intensity), measured in decibels, is the amount of energy present in a sound or sequence of sounds. This feature is also more complex than it may appear, as certain vowels have a greater intensity than others: a speaker may increase the air pressure when speaking if they are angry, or taking exercise, and may turn their head away, etc. Amplitude is still often found to be an accurate correlate for the perception of stress (Kochanski et al., 2005), despite all of these potentially complicating factors, which illustrates the way in which listeners are able to compensate for extraneous noise during the reconstruction of language. Formant structure refers to the range of frequencies which result from the resonance of sound within the vocal tract. F0 is the lowest formant (periodic wave) in the speech signal, and gives us the perception of pitch, and all the frequencies combine to give us the perception of timbre.

Although many authors are not concerned with the difference between the acoustic correlates and their perceptual cues, if we are to carry out research on how various prosodic features are perceived with a view to helping learners, then this difference is important. In the 1950s, newly available technology for acoustic analysis of the voice enabled Dwight Bolinger to formulate his pitch theory of accent (1958). Whereas previous experts had generally cited loudness as being the main cue to stress, (Bloomfield, 1933: 110), he was able to demonstrate the importance of pitch, by measuring and manipulating the physical correlate of F0. Dennis Fry's pioneering work at the Haskins laboratories using the Pattern Playback speech synthesiser to modify the acoustic parameters of word stress on minimal pairs differentiated by word stress (Fry, 1955; 1958; 1965) demonstrated the relative importance of the different physical correlates of the cues to the perception of stress in American English. Fry's work has encouraged many researchers over the years to investigate to what extent the four correlates of loudness, pitch, length and timbre contribute to the detection of prosodic cues such as stress, including myself (Frost, 2011); we will discuss the implications of some of this work in chapter four.

---

<sup>6</sup> I will use slanted brackets (/ /) for broad phonetic or phonemic transcriptions, and square brackets ([ ]) for narrow phonetic transcriptions, when giving extra details, such as allophones, etc.

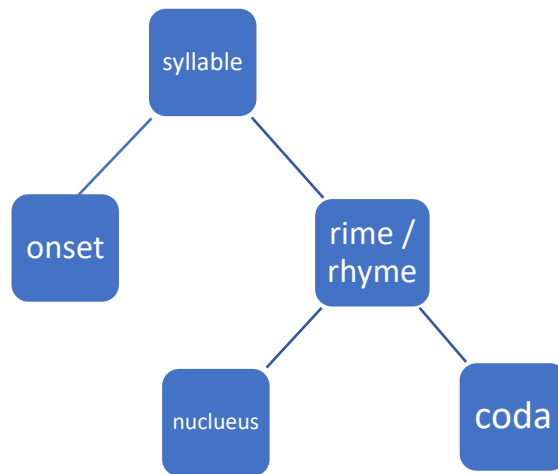
### 1.1.2. Stress

Stress refers to the degree of force used in producing a syllable (Crystal, 2008: 454). A stressed syllable is more stressed than an unstressed syllable, so stress is a relative phenomenon, or as Couper-Kuhlen explains, stress is “nothing more than the fact that in a succession of spoken syllables or words some will be perceived as more salient or prominent than others” (Cooper-Kuhlen, 1985: 19). We often distinguish the concept of linguistic stress from the phonetic and measurable reality of stress, also known as prominence (Cruttenden, 1997: 10). Sometimes stress is also referred to as “accent” (Bolinger, 1958), however to avoid confusion or ambiguity, I will usually refer simply to “stress”; if the need arises to differentiate between stress at the abstract level and at the level of production, and the context does not make it clear, I will use the terms “linguistic stress” and “prominence” respectively.

Stress is marked by the four acoustic cues described in the previous section, and their relative importance varies from language to language, from speaker to speaker, and also over time in a given section of discourse. In English, according to Bolinger’s theory of pitch accent (1958), the most important cue is often F0/pitch. As Jones explains in the preface to his *Outline of English Phonetics* (1918), this cue is often sufficient in English to mark stress: “in innumerable cases the requisite change in the direction of intonation without any increase of force whatever is sufficient to produce on the ear the effect commonly described as stress” (Jones, 1918: V). But perhaps the most problematic of the four cues to stress in English, is that of the formant structure of the phonemes, perceived as the timbre of vowels. As Cruttenden (1997: 19) states “unstressed syllables are usually shorter and often involve reduced vowels” and although we can distinguish up to seven levels (or “degrees”) of stress in English, they do not all have a distinguishing value. In fact, William Cooper & Stephen Eady (1986: 383) note that “there appears to be no upper bound on the number of degrees of stress distinguishable in a language”, so the degrees of stress various linguists have discerned in English are opinions that will vary from one author to another. From Otto Jespersen (1933) to Noam Chomsky and Morris Halle (1968), most authors accept that four levels of stress are sufficient for explaining meter. The degree or level of stress does have repercussions on segmental and suprasegmental features. To better understand stress, for the purposes of research and of teaching, it is necessary to distinguish three types of stress: **lexical stress**, **sentence stress** and **contrastive stress**. These functional categories of stress all share one important point – they are an attempt to focus the listener’s attention on the new and/or important information, be it at the level of a word, or a longer prosodic unit.

To discuss phenomena related to prosody, and especially stress and rhythm, we often talk of “stressed syllables” and “unstressed syllables”, but to be more precise, we need to refer to different parts of the

syllable. A syllable must be composed of at least one phoneme, and is usually composed in English of a series of consonants (C) and vowels (V): CV, CVC, CCVCC, etc. (Aslam & Aadil, 2011: 60).



**Figure 1.1.2a. Syllable structure, adapted from (Aslam & Aadil, 2011: 61)**

Figure 1.1.2 represents the traditional way of expressing syllable structure in English (Aslam & Aadil, 2011: 61). A CVC syllable, for example the word *cat* (/kæt/) is composed of an **onset** /k/ and a **rime**, or **rhyme**, which is made up of the **nucleus** (the vowel /æ/, and the **coda** /t/). The term **mora** is also used in describing some languages, such as Japanese, where it refers to a unit of time which has a structuring property in the language: it is the length of a short vowel, and a long vowel is two morae / moras. Stress generally affects the whole syllable, with more energy being detectable both in the onset, and in the rhyme / rime, especially on the nucleus, where the changes to F0, intensity, duration and formant structure are all measurable: for this reason, we generally talk of “stressed syllables”.

**Lexical stress** (also known as **word stress**) is a distinctive feature in certain languages, such as English, but also Spanish, Russian and Greek (Cruttenden, 1997: 15). In English, we differentiate for example between the verb *permit* /pə'mit/ and the noun *permit* /'pɜ:mɪt/<sup>7</sup> by using word stress. In longer words, the various levels of stress may have a bearing on the behaviour of syllables, for example syllables bearing secondary stress are rarely reduced. The patterns of where lexical stress occurs in English were mentioned in some of the earliest texts and pronouncing dictionaries of English (e.g. Walker, 1791), and were most comprehensively laid out by Chomsky and Halle (1968). My research has shown that even A1-A2 level French learners (according to the *Common European Framework of Reference for Languages*) nearly always place the primary lexical stress on the correct syllable when

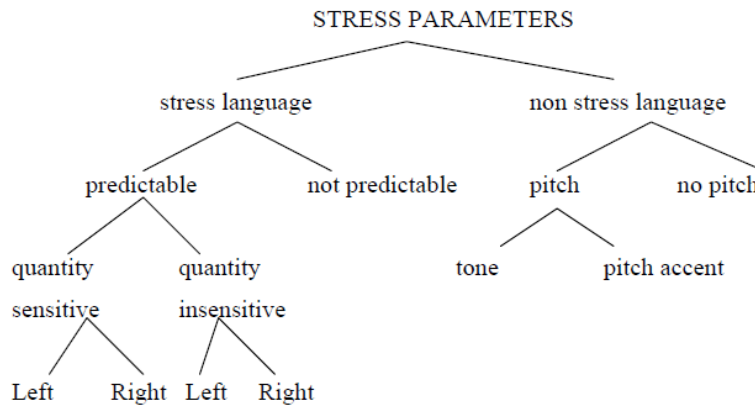
<sup>7</sup> Unless stated otherwise, I use British English in transcriptions.

speaking, even if they don't always produce all of the acoustic correlates to stress correctly (Frost & O'Donnell, 2018), and they rarely differentiate between levels of stress until C1.

**Sentence stress**, known variously as **prominence** (Celce-Murcia et al., 1996), **focus** (Cruttenden, 1997), **nucleus** (Wells, 2006), **pitch accent** (Pierrehumbert, 1980), and sometimes **primary stress** (Celce Murcia et al, 1996) is a similar phenomenon, but usually occurs over a larger prosodic unit than one word, generally referred to as a **tone unit** (TU) or a **tone group**. Tone units are usually a series of syllables, separated by pauses and containing a continuous intonation pattern, or series of pitches. One syllable in particular is the most prominent, known as the **nucleus** or the **head**. The syllables preceding the nucleus are known as the **pre-head**, and the syllables which follow are called the **tail**. Tone units generally (but not always) correspond to clauses, and the pauses often correspond to punctuation marks. Sentence stress in English usually falls on the last lexical word of the TU (noun, verb, adjective or adverb), but also falls on certain question words, proper nouns, etc. Sentence stress focuses the listener's attention on some of the most important elements in each tone unit, and as such is a crucial part of oral communication, LX users tend also to use more pauses, i.e. make smaller tone units, and it seems that these smaller tone units are linked more to meaning, than to syntactic structures, or at least they are more linked to the number of words for LX users, than to the number of syllables or to the syntactic structures involved (Volín, 2019). This non-typical phrasing of accented speech increases the cognitive load when it comes to processing the speech of LX users; as Sadi Phillips and her colleagues (Phillips et al., 2022) show, pauses in the right place makes processing easier for L1 users processing LX speech. Even in L1 speech perception experiments, unexpectedly long pauses interfere with speech comprehension (Martin, 1968).

Finally, **contrastive stress** is a form of sentence stress, but it can be placed on any syllable the speaker wishes to emphasise the particular syllable, or more commonly, word in which that syllable receives primary stress.

All languages can differentiate prosodically to some degree between syllables using contrastive stress, but not all languages have lexical stress. Heidi Altmann and Irene Vogel (Altmann & Vogel, 2002; Altmann, 2006), propose a stress typology model for stress languages (languages which have lexical stress) and non-stress languages (languages which do not have lexical stress).



**Figure 1.1.2b. Altmann & Vogel's Typology of stress parameters (Altmann, 2006: 31)**

The superimposed levels of stress and the complex interplay between the different forms and levels of stress, tone units and pauses, and rate of speech account for the effect of rhythm in a language.

### 1.1.3. *Rhythm*

Speech rhythm is difficult to quantify and measure, in fact some authors believe the concept of rhythm to be so difficult to pin down acoustically, it is better to think of it as a metaphor (Nolan & Jeon, 2014:9). Spontaneous human language is certainly not rhythmical in the same way as music, in that a measurable and regular alternation of strong and weak elements structures its progression through time in a predictable fashion over long stretches of discourse. It is, however, rhythmical in a “looser” sense, in that “its development in time is controlled by some hierarchical mental pattern giving each syllable a certain strength that controls aspects of its production, among which is its duration.” (Noteboom, 1997: 18). The fact that in English, the stressed syllable does not systematically occupy the same position within a word, as it does in Italian or Spanish, for example, together with the tendency of unstressed syllables to reduce in duration, means that the rhythms of English are complex. In order to describe the complex rhythms of English we may use the terms used to describe the metrical structure of poetry by the ancient Greeks, and still used today in the analysis of poetic metre. A line is composed of several “feet”, one foot being a stressed syllable (or beat) plus one or two unstressed syllables. It is also possible, though fairly unusual, for consecutive syllables to be stressed. The following patterns are the most common in English poetry (where an unstressed syllable is represented by “∪”, and a stressed syllable by “/”):

<b>iamb (iambus); iambic, as in</b>	<b><u>destroy</u></b>	/ ∪
<b>anapest (anapaest); anapestic</b>	<b><u>intervene</u></b>	∪ ∪ /
<b>trochee; trochaic</b>	<b><u>topsy</u></b>	∪ /
<b>dactyl; dactylic</b>	<b><u>merrily</u></b>	/ ∪ ∪
<b>spondee; spondaic</b>	<b><u>hum-drum</u></b>	/ /
<b>pyrrhic</b>	<b><u>the sea/son of/mists</u></b>	/ /

(Fussell, 1979: 20)

These six patterns, along with a few less common ones, combine over stretches of poems to form the metre of poetry. While it is true that spontaneous spoken language is rarely as metrically regular as poetry, containing as it does hesitations, repetitions, repair, etc., it may be described using these terms in many contexts.

Although these terms are used to describe poetry in many European languages past and present, the rhythmical patterns of European languages vary greatly. In some languages, lexical stress is fixed, such as Italian, where it usually falls on the penultimate syllable. In Greek, which gave us the terms above, lexical stress is largely dependent on morphology, and is marked with a written accent on the stressed syllable. In Scandinavian languages, although not marked with a written accent, lexical stress is also governed by morphology (Zora et al., 2016). In English, with its roots in Old Norse and Norman French, and with many words of Latin and Greek origins (O’Neil, 2019), the stress patterns are influenced by all of the source languages and the contact between them, and this accounts for the extremely complicated metrical structure of English.

#### 1.1.4. Isochrony

Isochrony refers to the timing of a language, i.e. to the relative weight (especially duration) of syllables, and to the rhythm which the relative weight of syllables confers to a language. The concept of isochrony has been argued over and contested ever since it was posited (Pike, 1945; Abercrombie, 1967). While it is almost impossible to prove acoustically in spontaneous speech as being a measurable physical reality, it nonetheless *seems* right, *feels* right, to anyone who is familiar with languages said to be on either side of the isochrony spectrum, as English and French are held to be. Pike refers to tone units as “rhythm units”, each one having one “strong stress”, and describes languages such as English



as having a “STRESS-TIMED rhythm” and languages such as Spanish as having a “SYLLABLE-TIMED rhythm” (Pike, 1945: 35). Of syllable-timed languages, he explains:

(...) it is the syllables, instead of the stresses, which tend to come at more-or-less evenly recurrent intervals--so that, as a result, phrases with extra syllables take proportionately more time, and syllables or vowels are less likely to be shortened and modified. (Pike, 1945: 35).

Ever since Pike first mentioned this stress-timed/syllable-timed distinction, isochrony has been argued and contested over by phoneticians and linguists. Abercrombie (1967: 92) confidently stated “As far as is known, every language in the world is spoken with one kind of rhythm or with the other” and that languages such as French, Telugu and Yoruba were syllable-timed, and languages such as English, Russian and Arabic were stress-timed. The isochrony debate, and the articles and studies it has produced, shed light on many of the problems which interest me as a researcher and teacher. In this section, we will explore some of the assertions and arguments relating to isochrony, since English and French are consistently mentioned as being representative of the two types of language in this dichotomy.

In his meta-analysis of the data on isochrony, Bertinetto (1989) concluded that stress-timed languages such as English have 6 properties:

- I. more intrasyllabic compensation;
- II. more [compensatory shortening] at the foot (and word) level;
- III. more vowel reduction in unstressed syllables;
- IV. more tolerance for extreme shortening of unstressed syllables;
- V. sharp contrast in the exploitation of prosodic features in stressed vs. unstressed syllables;
- VI. in general, less sensitivity to all linguistic and non-linguistic events localized on unstressed syllables.

(Bertinetto, 1989: 124).

Roach (1982) also found the tendency for stress-timed languages to reduce unstressed syllables to be important, though he was much more sceptical about the dichotomy than Bertinetto. He examined data from measuring inter-stress intervals for the six languages Abercrombie mentioned, and concludes that these measures do not allow for the categorisation of these languages according to the isochrony hypothesis. He does, however, admit that to examine the question thoroughly would entail a more detailed study, which would be problematic for several reasons, not least of which one would have to agree on tone unit boundaries, which is difficult. Roach (1982:78) believes that intra-language variation and inter-speaker variation account for bigger differences than being either stress-timed or syllable-timed languages, but he does not rule out the hypothesis altogether. He seems to see it as inaccurately framed, stating:

It would be necessary to consider possibilities such as that languages classed as syllable-timed may tend to have simpler syllable structure (Smith (1976) suggests this as a factor in the case of Japanese and of French), and that languages classed as stress-timed may be more likely to exhibit vowel reduction in unstressed syllables. (Roach, 1982: 76)

English does reduce its unstressed vowels, and tolerates extreme shortening of unstressed syllables, which Bertinetto refers to as “compensatory shortening” (Bertinetto, 1989: 106). However, can Bertinetto’s six points be borne out by measuring speech across corpora of spoken languages to prove the isochrony hypothesis? Several authors have taken acoustic measurements to attempt to prove or disprove the dichotomy of the isochrony hypothesis. Dauer (1983) compared interstress intervals in five languages, and found that interstress intervals were no more regular in English, a stress-timed language, than in Spanish, a syllable-timed language. She therefore concluded that the hypothesis was invalid, and that languages form a rhythm continuum from least to most stressed-based. She also concluded that it is actually very difficult to decide on what would be acceptable measurements to validate stress-timing and syllable-timing as an acoustic reality. Andy Arleo (1995) came to a similar conclusion from studying nursery rhymes and skipping rhymes in English and French, concluding that there is a scale of isochrony across speech, ranging from arrhythmic spontaneous speech, to more isochronous genres, such as cheers, nursery rhymes, chants, poetry, etc. Antonio Bertrán (1999) measured the absolute duration of feet from the onset of the stressed vowel until the next stressed vowel in seven languages, and obtained results that “openly contradict the typological models they are supposed to represent.” (Bertrán, 1999: 125). And yet there are clearly differences between the rhythmic structures of languages, and it is understandable that phoneticians would attempt to classify languages according to their rhythmical structure.

Most authors agree that attempts to classify all the world’s languages into two groups based on their rhythm is futile. As Ilse Lehiste (1977) concludes, “Most investigators have therefore either rejected the claim that English is a language characterized by isochrony, or have attempted to reinterpret the experimental findings to take into account the fact that perfect isochrony cannot be found in production.” (Lehiste, 1977: 255). Marina Nespór (1990) was one of those researchers who attempted to redefine the categories of isochrony: she argued against traditional rhythmic categories and proposed the existence of “intermediate languages”, which, she maintained, exhibit some properties associated with stress-timing and some associated with syllable-timing. Following this presumption, Franck Ramus and his colleagues (1999) studied eight languages including English and French. They measured the proportion of vocalic intervals (%V) and the variability of consonantal intervals ( $\Delta C$ ), and conducted a series of task-based experiments, including with infants. They found that “rhythm contrasts are accounted for by differences in the variety of syllable structures” (Ramus et al., 1999: 274), and they remain “agnostic” about the possibility of being able to class all the world’s languages

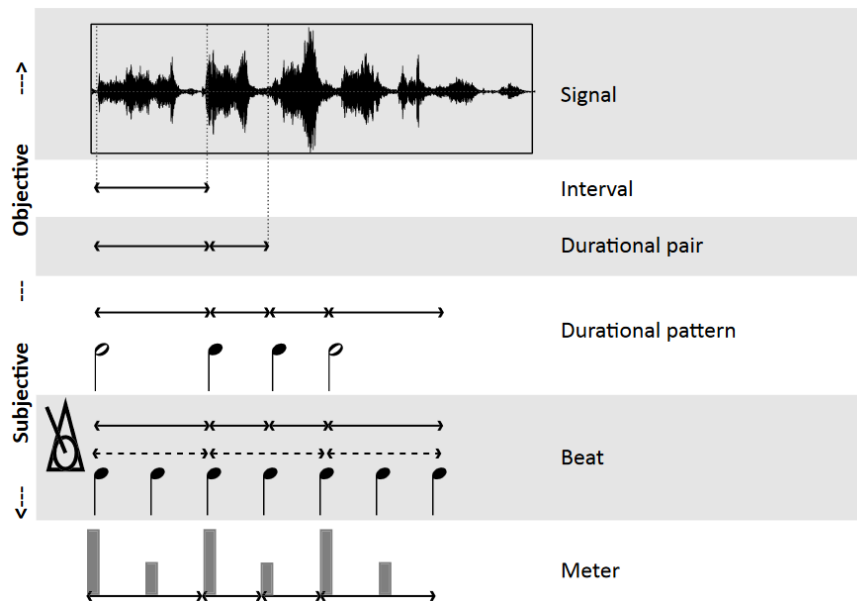
into three categories, stress-timed, syllable-timed and intermediate, suggesting that such a simplistic categorisation would certainly break down as more languages were added to the eight they studied. But if we limit ourselves to the differences between English and French, and do not attempt to classify *all* the world's languages, we do find some clear differences: Ramus and colleagues found that %V was smaller in English than in French, which does not have vowel reduction. On the other hand,  $\Delta C$  was larger in English and reflected the more complex syllable options available in that language. So, although studies do not support the isochrony hypothesis for all languages studied, there are important metrical differences between French and English.

The idea of a spectrum, from languages where stress is strongly marked and unstressed syllables greatly reduce, to languages where syllables are more rhythmically similar, is taken up by several authors. As Cruttenden (1997) states: "All the evidence suggests that both stresses and the number of syllables influence rhythm in all languages but particular languages have a tendency to give greater or lesser weight to the stress factor." (Cruttenden, 1997: 21). Arvaniti (2009) also believes that a relative, rather than an absolute approach is more useful when trying to understand language rhythm, in that measurements would need to be adapted to the prosody of each language under consideration (Arvaniti, 2009: 61) but even then, certain acoustic measurements would have to be ignored.

If we are to consider isochrony as existing at all, then it seems to exist on a spectrum, rather than as a dichotomy. Furthermore, the same measurements cannot be applied to different languages, as the very features which contribute to stress and therefore rhythm, are language-dependent. And, of course language-specificity is as true of perception as it is of production. Donia Scott and her colleagues (1985) conducted two experiments to explore the rhythmical differences between English and French. Interestingly, they were unable to prove any regularisation which would help support the stress-timed versus syllable-timed distinction, much less the isochrony principle. However, Scott and colleagues' finger-tapping experiment produced different performances between English and French speakers, indicating that listeners perceive language differently, specifically they group syllables differently. Jakobsen, Fant and Halle (1951) illustrate the influence of a language user's own language when perceiving an LX by the way different L1 users react to the same stimuli based on the rhythmic patterns of their own languages. In a series of experiments, speakers of different L1s were asked to listen to a series of knocks at regular intervals, with every third knock louder. Jakobsen and his colleagues report:

The pause is usually claimed by a Czech to fall before the louder knock, by a Frenchman to fall after the louder; while a Pole hears the pause one knock after the louder. The different perceptions correspond exactly to the position of the word stress in the languages involved: in Czech the stress is on the initial syllable, in French, on the final and in Polish, on the penult. When the knocks are produced with equal loudness but with a longer interval after every third, the Czech attributes greater loudness to the first knock, the Pole, to the second, and the Frenchman, to the third. (Jakobson, Fant & Halle, 1951: 10-11).

As we have already seen, speech contains measurable acoustic parameters and perceptual cues. Humans are particularly good at extracting metrical information from an audible signal (Iverson et al., 2006, Patel, 2006) and they reconstruct subjective notions such as meter from this information.



**Figure 1.1.4. Levels of Temporal and Rhythmic Structuring (Kotz et al. 2018: 898)**

In Figure 1.1.4, Sonja Kotz and her colleagues illustrate the way listeners extract a pulse from an audible signal and construct subjective metre. This process, which as will see in the next chapter is central to how humans acquire language and how human language developed, partially explains why the notion of isochrony persists. Cauldwell (1996) also believes the “so-called stress-timing of English” to be more closely linked to perception than to production, concluding in his article on stress-timing:

If patches of stress-timing do occur this is an incidental, patchy effect brought about by fleeting coincidences between time and the occurrence of prominences and word-accents. Speakers time language: language does not time speakers. (Cauldwell, 1996: 10).

Cauldwell does, however, concede that much more regular prototypical speech occurs in non-spontaneous speech, such as nursery-rhymes. Gibbon (2018), also places store by perception in matters of rhythm in a paper examining models based on corpora from different languages examines isochrony and the concept of rhythm in language, specifically by referring to five different “time domains” (discourse time, individual time, social time, historical time, and evolutionary time), and two opposing concepts of time (“clock time” and “rubber time” – the first being measurable, and the second being perceptual and relative). Even by redefining time, as he states in his conclusion: “Neither all nor only empirically observable rhythms can be accounted for” (Gibbon, 2018: 8). Finally, however, he concedes, as does Cauldwell (1996), that the rhythm of English is much more regular when it comes to formulaic language, such as idioms, collocations, etc., especially regarding the placement of pauses,

and therefore the size and nature of tone units (Lin, 2012), which is a concept I will touch on again in chapter two, and in the final chapter when dealing with prosody in learning English.

In conclusion, I would like to make four points. Firstly, it seems that a dichotomous isochrony model which attempts to classify all the world's languages into either stress-timed or syllable-timed languages, does not stand up to analysis when acoustic realities in spontaneous speech are measured. And yet the impression of two categories persists. From perceptual experiments cited above, and bearing in mind that the measurable acoustic cues to stress have their perceptual cues, there is something in people's perception of metrical structure in language which leads them to hear different rhythmic structure. It is reasonable to conclude that isochrony is therefore at least partially a perceptual phenomenon. Secondly, the dichotomous theory of isochrony does not bear up to analysis, but there does appear to be a spectrum (acoustic and/or perceptual), from more stress-timed to less stress-timed languages. Thirdly, spontaneous speech is not the only speech we use in our lives, and therefore not the only speech use for acoustic measurements. Infant-directed speech, nursery rhymes, poems, proverbs, idioms, insults and other formulaic speech are much more prosodically rich than spontaneous speech, and are more rhythmically regular – and they are a large part of how we learn our first language or languages. If we take such examples of speech as being close to a prototype, or a sort of idealised version of the metrical structure of a given language, we may be closer to the mental representation language users have of their own language, and to the filter through which they are likely to perceive other languages. Finally, it is also possible to reframe the debate in relative, rather than in absolute terms, as Arvaniti (2009) proposes. If we compare English and French, they *do* have very different metrical properties. Firstly, English has lexical stress, French does not. Secondly, as As Wenk and Wioland rightly pointed out, French is “trailer-timed” and English is “leader-timed”. And finally, as a language which marks stress strongly, English systematically reduces its unstressed syllables, often considerably, as Bertinetto (1989) noted, whereas French does not to the same degree. It is unsurprising therefore that, as several studies have found (Ramus et al., 2000), there are very important differences in the rhythmical structure between French and English, often held up as archetypes of these two categories.

These differences between French and English are central to many of the problems French learners have with English, as we shall see in chapter three.

#### *1.1.5. Intonation*

The term “intonation” has a variety of meanings, ranging from narrow to broad definitions in its scope. Intonation, in its broadest sense, is the melody of the language. There is a slight difference, as I set out

at the beginning of this chapter, between the American and the British schools when it comes to analysing intonation, although this difference is somewhat reduced today. Traditionally, American linguists used the term “tone” to refer to word-level pitch patterns and “intonation” to refer to changes at the level of the utterance or the sentence (Pennington & Rogerson-Revell, 2019: 5). I will use the term “intonation” or “intonation contours” to refer to “the use of suprasegmental phonetic features to convey ‘postlexical’ or sentence-level pragmatic meanings in a linguistically structured way” (Ladd, 1996:4).

Intonation, although it is sometimes used interchangeably with “tone”, “pitch”, and other terms, is when a speaker uses a variety of suprasegmental phonetic features (length, loudness and pitch – but also pauses, and other phenomena), over sequences of speech which may be longer than single words. As t’Hart & Collier have pointed out, intonation may be analysed on three levels, depending on the degree of abstraction. At the acoustic level, the human ear cannot accurately perceive (or rather selectively perceives) these phenomena, so we can distinguish a second level of perceivable pitch events at the phonetic level. Finally, at a more abstract level, we can group these pitch events into meaningful categories based on their function. These three levels are particularly important for discussing intonation, as they are for teaching it, since it is often very difficult to identify intonation patterns, distinguish between rising and falling tones, etc. This has been borne out by several studies, which showed up to a 50% failure rate at identifying consciously contours taken out of context (Hadding-Koch, & Studdart-Kennedy, 1964), and by my own experience in Albert Di Cristo’s Master’s classes in experimental phonetics, and by my experience teaching intonation.

As this report is particularly concerned with English, I will briefly summarise the different intonation patterns commonly identified for English. Firstly, Pike (1945) and Trager and colleagues. (1951) both identify four tones: low, middle, high and extra-high. Secondly, intonation may do four basic things: rise, fall, fall-rise or rise-fall (Roach 1991). Finally, we can identify a certain number of contour patterns for actual speech: Cruttenden (1986) identified seven canonical intonation patterns in British English, which he calls “nuclear tones” (Cruttenden, 1997: 54). These tones are low fall, high fall, rise-fall, low rise, high rise, fall-rise, and mid-level, and this classification is sufficient for most teaching and indeed research needs. In recent years, it has proven necessary to add the High Rising Terminal (HRT), or “upspeak” / “uptalk” (Wells, 2006: 37). This pattern may have started in New Zealand, or Australia, but it has spread rapidly to British English from American English to such an extent in the last 20 years or so, that it is almost ubiquitous in the speech of thirty-somethings and below (Warren, 2016). The addition of the HRT contour has been picked up on by many French learners, but often manifests as a step-up on the last syllable of the rhythmic group, rather than as a progressive rise over the syllables forming the tail.

### 1.1.6. *The functions of prosody*

The main function of prosody is conveying semantic, pragmatic, and syntactic information. But most models of lexical processing suggest stages, from initial processing of the acoustic signal to comprehension (Rost, 2011), and before a listener can even access this information, speech must be segmented into meaningful chunks. We do not often speak with pauses between words, and prosody is evidently important in this process of segmentation: initially, in what we may call linguistic processing, in, and later on, in processing at the semantic and pragmatic levels. As Cutler (1995) summarises:

The evidence suggests that prosodic information will participate in the initial activation stage of word recognition to the degree that its use is efficient. If the point at which prosodic information becomes useful is too late (because the word candidates it might have contributed to activating are already activated), or if it contributes no added value to the activation code over and above that which is contributed by segmental information, then the word-recognition process will not benefit from it. (Cutler, 1995: 185)

Central to the work of a language teacher and a researcher interested in oral language, prosodic cues therefore, help us to process spoken language, to segment and parse utterances, to separate segments of fluent speech into meaningful chunks, such as phrases, and lexical items (Cutler et al., 1997). An example of this in English is stress-demarking intonation groups (tone units), with strong initial and stress and final lengthening. Also, unaccented syllables may be pronounced more rapidly at the beginnings of intonational phrases beginning with *and then be-* or *that if we*, etc. (Cruttenden, 2014: 278). Certainly, prosody is not the only factor which helps listeners to segment words: it is in a certain sense in competition with other factors, including pauses and gesture. However, in normal speech, the four prosodic parameters mentioned in the introduction play an essential role in the segmentation of words, crucially allowing us, in subsequent stages of processing, to call upon semantic, grammatical, and pragmatic information to attribute meaning to utterances.

Wells (2003) lists the following functions of intonation: attitudinal, grammatical (or syntactic), focusing (accentual or informational), discourse (or cohesive), psychological, and indexical (p. 10-11). The intonation patterns mentioned in section 1.1.5 are associated with various functions, as Cruttenden (2014) explains:

In general, falling nuclear tones (whether  $\downarrow$ low fall,  $\uparrow$ high fall, or  $\wedge$ rise-fall) are separative, matter-of-fact and assertive; whereas  $\downarrow$ low rise,  $\uparrow$ high rise and  $\sim$ fall-rise are continuative, implicative and non-assertive. Level tones (most common among these being the mid level) belong with the rising tones in the sorts of meanings they convey. (Cruttenden, 2014: 292).

Certain functions of intonation are present across languages, such as a rising tone signalling incompleteness or a request for information, and a falling tone signalling finality, or an assertion. This

phenomenon is often thought of as universal, or at least as being common across a wide range of human languages. Ultan (1978) suggested that:

the widespread contrast between a terminal falling and a terminal rising contour representing a meaningful distinction between an attitude of finality or conclusiveness and one of suspension, incompleteness, doubt, questioning, or the like on the part of the speaker. (Ultan, 1978: 45)

Ultan analysed the interrogative systems of 53 languages, and on the subject of prosody, he drew three conclusions for intonation and two for stress. Regarding intonation, yes-no questions consisting of rising terminal, higher pitched, or special stress contours can be found in nearly all languages. Secondly, tag questions with non-rising (or higher pitched or stressed) contours implies also tag questions with rising contours. And finally, he found is a considerably better than chance probability that an intonation type consisting of rising terminal, higher pitched, or special stress contour may occur in languages of all basic order types. Regarding stress, he found a slight tendency for question particles to occur with higher pitch or prominent stress in SOV languages, and that question words tend to occur with higher pitch or prominent stress in languages of all basic order types. (Ultan, 1978: 54-55). Ultan accepts that his sample is not exhaustive, and refrains from using the word “universal”, preferring words like “common”, “general” and even “near-universal” (Ultan, 1978: 51), as certain stress and intonation patterns can be found across most of the languages he studied. And of course, prosody is also used to signal other syntactic information, such as the difference between a defining or a non-defining relative clause, or pragmatic functions such as listing, etc. where rising intonation also signals incompleteness. Although many of these prosodic features are similar across languages, the exact nature of each feature in different languages is often different, and this is where the problems occur for language learners; as we shall see in the next chapter: a rising or falling intonation pattern may signal the same basic function, but not in exactly the same way.

Prosody may also convey information on cognitive and affective states: certain emotions have been shown to have an effect on prosody, for instance vowel length (anger and sadness), high intonation (joy) and flat intonation (complaint) (Fónagy & Magdics, 1963). And of course, no utterance exists without a context, and although a rising intonation pattern may signal incompleteness, it may also mean that the speaker is thinking about something, or it may convey humour, irony, sarcasm, etc. (Attardo et al, 2013). As Pike (1945) explains, intonation may convey meaning in a particularly powerful way, more powerful even than the words a speaker uses:

An extraordinary characteristic of intonation contours is the tremendous connotative power of their elusive meanings. One might hastily and erroneously assume that forms which change so rapidly and automatically could not be semantically potent. Actually, we often react more violently to the intonational meanings than to the lexical ones; if a man's tone of voice belies his words, we immediately assume that the intonation more faithfully reflects his true linguistic intentions. (Pike, 1945: 22)



Prosody conveys much more than linguistic information, it also fulfils other roles, helping us for example to identify and convey emotions:

Sometimes against our will, it [prosody] signals or helps signal information about our sex, our age, and our emotional state, as part of a parallel communicative channel that can be interpreted by listeners (even some non-human ones) who do not understand the linguistic message. (Ladd, 1996: 1)

Finally, an ability to perceive suprasegmental features has been linked to other cognitive, social and emotional traits, such as empathy (Rota and Reiterer, 2009), and it has long been known that many people with autism and Asperger's have difficulty identifying prosodic cues (Cleland & Peppé, 2003). While working on the *Innovalangues* project, and in an attempt to measure learner's prosodic skills, we searched for tools which could do this, and although we found nothing specifically aimed at language learners, we did find a tool designed for English L1 users with autism and spectrum disorders, *Profiling Elements of Prosody in Speech-Communication* (PEPS-C) (Peppé, 2013). We contacted Sue Peppé in 2014, and purchased her test kit, which included a program, and a small user interface, with clearly marked buttons for answering the questions in the test. The test itself comprises 14 tasks, and addresses six linguistic functions, such as identifying turn-taking, affect, lexical stress, phrase stress, focus, etc. We tested the kit on several English LX learners from Grenoble Alpes University, and although the principle of the tests could have been adapted to our needs, it was not suitable for our purposes. Ultimately, we chose a different approach to measure prosody, which I shall describe in the final chapter of this project, and we did not publish the results of this pilot study. However, the exchanges with Sue Peppé, and the process of running the tests taught me a great deal about the nature of prosody and how to construct activities to measure it. Having defined some of the key terms in prosody, in the next chapter, we shall explore some of the many different functions of prosody.

## CHAPTER TWO

### First language prosody

In this chapter, we will begin by looking at the physiological, neurological, and cognitive issues surrounding prosody, then review the way prosody has shaped human language, before finally examining individual and societal issues concerning prosody.

#### 2.1. The physiological components of prosody

The four acoustic correlates to prosody outlined in the previous chapter (F0, duration, amplitude, and formant structure) function as cues which allow us as listeners to distinguish sounds, to segment them into meaningful chunks for processing, and to access the information they convey. In face to face speech at least, these four correlates are always accompanied by other physical signs, such the necessary movement of articulators, facial expressions, and hand gestures. And when we use prosodic cues to highlight information in speech, we also use more physical effort in some way or ways.

It is customary to think of the mind as being seated in the brain, and of course, the brain is indeed where most of our thinking and learning takes place. But the brain is also part of our body, and the way we move, breathe, and interact with our environment all have an effect on our cognitive and affective processes – and, of course, our language. Lakoff and Johnson highlight the fact that we actually *need* our bodies – and not just a disembodied mind – to reason, stating: “[t]he same neural and cognitive mechanisms that allow us to perceive and move around also create our conceptual systems and modes of reason” (Lakoff and Johnson, 1999: 4). The activity of the mind, as that of the physical envelope which houses it, takes place in time and space, or as Andy Clark (1999) states “the activity of an essentially *situated* brain: a brain at home in its proper bodily, cultural and environmental niche” (Clark, 1999:1). Far from being a controversial field, it has been widely accepted for decades now that this is what Margaret Wilson (2002: 635) refers to as “a very general underlying principle of cognition”. This conception of the mind, the body and the environment as an integrated system has become known as “embodied cognition” (Varela et al., 1992), or “grounded cognition” (Barsalou, 2008). As Wilson and Golonka (2013) state:

Our bodies and their perceptually guided motions through the world do much of the work required to achieve our goals, replacing the need for complex internal mental representations. This simple fact utterly changes our idea of what ‘cognition’ involves, and thus embodiment is not simply another factor acting on an otherwise disembodied cognitive processes. (Wilson & Golonka, 2013: 1)

So, let us begin not in the brain, but with the physical processes responsible for producing prosodic cues.

### *2.1.1. The vocal apparatus & prosody*

The Chambers dictionary of English defines “voice” as “[s]ound produced by the vocal organs of living beings, *esp.* of human beings in speech or song [...] sound uttered with resonance of the vocal chords (phonetics)” (Chambers, 1993: 1946). Speech is produced by air pressure passing through the vocal cords, and the sounds made by the vibrations in the vocal cords are transformed by the speech organs in the vocal tract, from the glottis up to the lips. When teaching pronunciation, most teachers concentrate on the articulators, such as the lips, the tongue, and the teeth, and these visible speech organs are primarily responsible for the realisation of phonemes. Jones (1918: V) observed: “stress is generally accompanied by a change in the direction of intonation, and that this change in the direction of intonation is of greater importance than any increase in the force of the breath”, and it is in the lungs that this increase in pressure is produced. When we inhale during normal breathing, there are two mechanisms which work together, the diaphragm and the intercostal muscles, but it is primarily the diaphragm which lifts the ribcage, inflating the lungs (Tortora & Derrickson, 2009: 890-893). So-called shallow breathing, or costal breathing is mainly inhalation using the intercostal muscles, and deep, or abdominal breathing is, in fact, diaphragmatic breathing, although we feel our stomach moving, and may voluntarily project our stomachs outwards when doing this sort of breathing. Expiration, or breathing out, is usually gravity plus the natural elasticity of our muscles and membranes allowing the lungs to return to rest, thus expelling air. We can accelerate this procedure by mobilising our abdominal muscles to aid our diaphragm to expel air more quickly, as we do when we laugh, sigh, and sob (Tortora & Derrickson, 2009: 894). This mechanism - i.e. our diaphragm working a little more than it normally would, and assisted slightly by our abdominal muscles - is how we achieve the extra sub-glottal pressure required to produce the acoustic cues associated with stress, particularly increased intensity. However, an increase in pressure does not simply result in an increase in the cue of amplitude, and allow us to perceive the nucleus of a syllable, i.e. the vowel sound, as louder, it also increases F<sub>0</sub>. We generally increase F<sub>0</sub> by increasing the tension of our vocal cords by contracting the cricothyroid muscle (Tortora & Derrickson, 2009: 881), but all other things being equal, if the flow of air from the lungs through the vocal cords is increased, the vocal cords will also vibrate more rapidly. Indeed, if we attempt consciously to raise the intensity of a vowel sound in a syllable, F<sub>0</sub> will often increase too, just as if we try to increase F<sub>0</sub>, amplitude will increase too. Performing either of these acts will also generally increase the duration of the syllable, so closely linked are these three prosodic cues: breathing and phonation mechanisms act together to produce prosodic cues.

Apart from the phonation involved in producing prosodic cues, there are also facial and physical gestures which very often act as cues, either voluntary or involuntary, creating what McNeill (1992) refers to as a cohesive function. The “McGurk effect” first demonstrated in a study by Harry McGurk and John MacDonald (1976), showed that listeners could be fooled into hearing phonemes which correspond to what they see, and not what is actually said. Interestingly, this effect was much more pronounced in adult listeners (92%) than in children (59%), showing the primacy of visual cues over auditory cues is even stronger in adults (McGurk & Macdonald, 1976: 747), a finding which has important implications for pronunciation instruction with older learners.

The hands, in particular, very often accompany focus (Guellai et al., 2014: 127). The relationship between facial gesture and physical effort has long been established (McKenzie, 1905), and so it is perhaps unsurprising that certain involuntary facial expressions such as eyebrow raises accompany the prosodic cues which mark focus, for example. Dwight Bolinger referred to the use of head movements and gesture when marking sentence stress in English as “this pugilistic obligato to the linguistic tune” (1958: 127). One of the roles of prosody, as we saw in the last chapter, is to convey attitude and emotion, so it is also quite normal that facial expression should accompany this dimension of prosody too, for example furrowing eyebrows with an extra-high falling contour to express disbelief, Cavé and his colleagues (Cavé et al., 1996) established correlations between eyebrow movements and F<sub>0</sub>, showing in fact that the left eyebrow correlated more strongly than the right with F<sub>0</sub> variations. These eyebrow raises are predictable and regular in English, as Gast’s study of late-night TV show interviews showed, occurring on average some 630 ms after the pitch contour peak (Gast, 2023).

Swerts and Kramer (2008) demonstrated that listeners were quicker and more successful at picking up on auditory prosodic cues when accompanied by appropriate facial gestures in English. Furthermore, they also found that the upper face is more important than the lower face. Dohen et al. (2009) found that French speakers also identify contrastive stress more quickly and accurately in French when visual cues are also given, and Rapin and Ménard (2019) obtained similar results for children in French, in a protocol using lip positioning as a cue, showing that even without the extra cues offered by the upper face, the extra movement, or hyperarticulation in the lips which accompany prosodic cues help children identify the syllables receiving contrastive focus. This close relationship between physical movements and meaning is the reason why prosody is the most iconic aspect of human language.

### *2.1.2. Iconicity*

By iconicity, I mean here the correspondence of form and meaning. A highly iconic sign in most cultures would be, for example pointing at one’s own chest to signify “I” or “me”. Most written and spoken

language is, therefore, not iconic, whereas many gestures can be iconic. Ferdinand de Saussure famously characterised human language as arbitrary signs:

*Le lien unissant le signifiant au signifié est arbitraire, ou encore, puisque nous entendons par signe le total résultant de l'association d'un signifiant à un signifié, nous pouvons dire plus simplement : le signe est arbitraire*

The sign which unites the signifier and the signified is arbitrary, or rather, since we understand the sign as being the result of the association of the signifier and the signified, we may simply state: the sign is arbitrary. (Saussure, 1916: 100).

By arbitrary – the opposite of iconic – Saussure means that there is no direct or necessary correspondence between form and meaning, and it is certainly true that most letters, phonemes, or even entire lexical items – and the meanings they convey - do indeed seem completely random, or arbitrary. However, occasionally, on the segmental level, the way certain phonemes and words sound may well be linked to meaning. Aside from onomatopoeia and certain exclamations, which Saussure excludes from his definition of the linguistic signs (Saussure, 1916: 101-102), some signs are not entirely arbitrary. Some word forms are not entirely arbitrary, as cross-linguistic studies have shown (Nygaard et al, 2009), but these are not frequent and often refer to spatial relationships.

Prosody however, differs from phonemes in this respect. As we saw earlier, certain aspects of intonation may be considered as language universals, so prevalent are they across different languages: especially the final fall signalling completedness and the final rise signalling uncompletedness (Ulтан, 1978). In many languages, intonation changes and head movements or other gestures coincide to indicate turn-taking (Stivers et al., 2009), for example an upwards nod and a corresponding rising intonation pattern indicating that the turn has passed to the co-interlocuter, and that is up to them to complete the missing information and / or continue the dialogue.

It is with stress, however, that the iconicity of prosody is most evident. The realisation of linguistic stress, be it at the word level or at the level of bigger prosodic units, involves highlighting a given segment of speech, with a number of physical features, some auditory, some not. In terms of the auditory cues to stress, the peak intensity and F0 usually affect the vowel which forms the nucleus of a syllable, as we saw in the previous chapter; visible cues to stress usually also correspond to this syllable, and sometimes extend beyond one syllable. When a syllable is stressed, there is increased pulmonary pressure, therefore more physical effort. This can be heard, and indeed measured, by the acoustic prosodic correlates to stress. However, these are not the only physically measurable changes a speaker uses to denote linguistic stress, to highlight new and/or important information. A speaker may also choose to point, bang the table, make other gestures either voluntary or involuntary (such as eyebrow-raising) as the syllable in question is realised. Alexandra Ćwiek and Susanne Fuchs (2019) showed that there is a statistically significant correspondence between head raises and rises in F0, an

effect which can be frequent in certain speakers, and that when talking about large objects, the head elevations are also larger (Ćwiek & Fuchs, 2019: 6-7). With stress, these physical and non-verbal signs coincide with meaning, which makes stress, more than with any other feature of human language, a highly iconic feature of human speech.

## 2.2. Prosody and cognitive neuroscience

Having explored the physical side of prosody, let us turn our attention to the way the brain processes and produces prosody. Of course, we do not have access to the brain and its workings directly. Huge improvements in data processing and the precision of neuroimaging techniques in the last two decades have enabled us to see with increasing precision *where* the chemical and electric activity takes place, and combined with behavioural experimentation, we are able to make comparisons and draw conclusions. It is not within the scope of this report to cover this subject in depth, so I will limit myself to a very brief presentation of the brain, before exploring how the brain processes prosody, both on a visual and auditory level.

The human nervous system makes up around 3% of body weight, and comprises the brain, the and a complex system of nerves linking the brain to sensory receptors. The adult brain is made up of of four major parts: brain stem, cerebellum, diencephalon, and cerebrum. The latter is the largest, and is divided into two hemispheres, each consisting of four lobes: the frontal, temporal, parietal, and occipital lobes (Tortora & Derrickson, 2009: 514). The different regions of the brain are linked to each other and to other parts of the head and body by 12 cranial nerve pairs. The brain's hemispheres and the lobes are associated with certain functions, for example executive functions are concentrated in the frontal lobe. Information circulates in the brain via neurons, of which the human central nervous system (the brain and spinal cord plus the peripheral nervous systems, including the gut) contains "approximately 100 million" (Tortora & Derrickson, 2009: 426). Neurons vary in size and shape, but are essentially made up of "grey matter", which is made up of soma and its dendrites. Dendrites are coated with "white matter", and at the end of each branch is a synaptic end bulb, which can connect with other bulbs, forming neural connections, or synapses, along which information can travel in the form of electrical signals (Tortora & Derrickson, 2009: 418). The average neuron in the cortex has 7000 synapses (Pakkenberg et al., 2003), and the more frequent the activity, the larger and more permanent the synapses. This reinforcing of synapses constitutes a physical - or rather biological – means of measuring learning.

Our understanding of the processing and production of language has greatly improved over the years since Richard Sperry's "split brain" theory of brain lateralization (Sperry, 1961). Sperry's experiments

on monkeys and cats led him to suggest that a “split” but connected brain made it possible for one hemisphere to concentrate on analytic procedures, while the other “spare” half could concentrate on other tasks (Sperry, 1961: 5). Some authors have claimed, principally after studying speech disorders, that linguistic processing, such as encoding or decoding syntactic and lexical information, occurs mainly in the left hemisphere of the brain (Beaumont, 2008). However, certain studies show that a similar amount of lexical processing takes place in each hemisphere (Hickok et al., 2008). As neuroimaging has improved, we have been able to visualise brain activity with greater accuracy, and we now have a clearer understanding that the brain is two interconnected hemispheres, which work together in all our cognitive activity (Sousa, 2017: 191).

At the end of the twentieth century and beginning of the twenty-first century, neuroimaging began to shed light on which regions of the brain are engaged in different aspects of language, but lacked precision (Fedorenko & Kanwisher, 2009; Price, 2012). Great strides have been made over the last twenty years as imaging technology and the ability to process big data have improved, using techniques such as Electroencephalography (EEG), electrocorticography (ECoG), and positron emission tomography (PET) to compliment functional magnetic resonance imaging (fMRI). Many advances have been made in our understanding of where language processing and production happens, for example it has been shown in recent studies that humans with more robust white matter tracts in the arcuate fasciculus exhibit better language learning abilities (Wong et al., 2011). These advances have led some neuroscientists to favour network-based approach to complement the mechanistic and lesion-based approach which necessarily preceded modern neuroimaging (Bertolero & Bassett, 2020). Network neuroscience uses two types of items to represent neural activity: nodes, i.e. regions of the brain such as neurons, or groups of neurons, and edges, i.e. structural connections, typically white matter or axons (Bertolero & Bassett, 2020: 1274). In this way, large amounts of data from many different sources can be represented in terms of networks using mathematics to help refine models, for example identifying hubs, where nodes can be shown to participate more frequently in establishing connections.

### *2.2.1. Visual and motor activity in the brain*

Linguists such as Daniel Jones have long supposed that we have an “inner voice” (Jones, 1918) and as we will see in the next chapter, the motor theories of perception of the 1950s and 1960s were based on this idea. It is widely understood that the brain is capable of “the mental simulation of external events” (Wilson, 2002: 635), and as neuroimaging has improved, the links between cognition and motor activity in the brain have been further explored by early work on apes and monkeys, and later

on humans. A group of Italian researchers (Rizzolatti et al., 1988) working on macaque monkeys were able to observe that the F5 area of the brain, analogous to our Broca's area, was excited by monkeys watching their handlers or other monkeys making gestures such as giving food. The same part of the brain was activated whether the monkeys made the gestures or observed the gestures: the monkeys were, in effect, imagining themselves making the same gesture, or receiving the food. They called this activity "motor evoked potentials" (MEPs). Rizzolatti and his colleagues continued their work and tested their hypotheses on humans (Fadiga et al., 1995), concluding that:

The excitability of the motor system increases when a subject observes an action performed by another individual. Furthermore, the pattern of muscle activation evoked by transcranial magnetic stimulation (TMS) during action observation is very similar to the pattern of muscle contraction present during the execution of the same action. These findings indicate that, in humans, there is a neural system matching action observation and execution. (Fadiga et al., 1995: 2609)

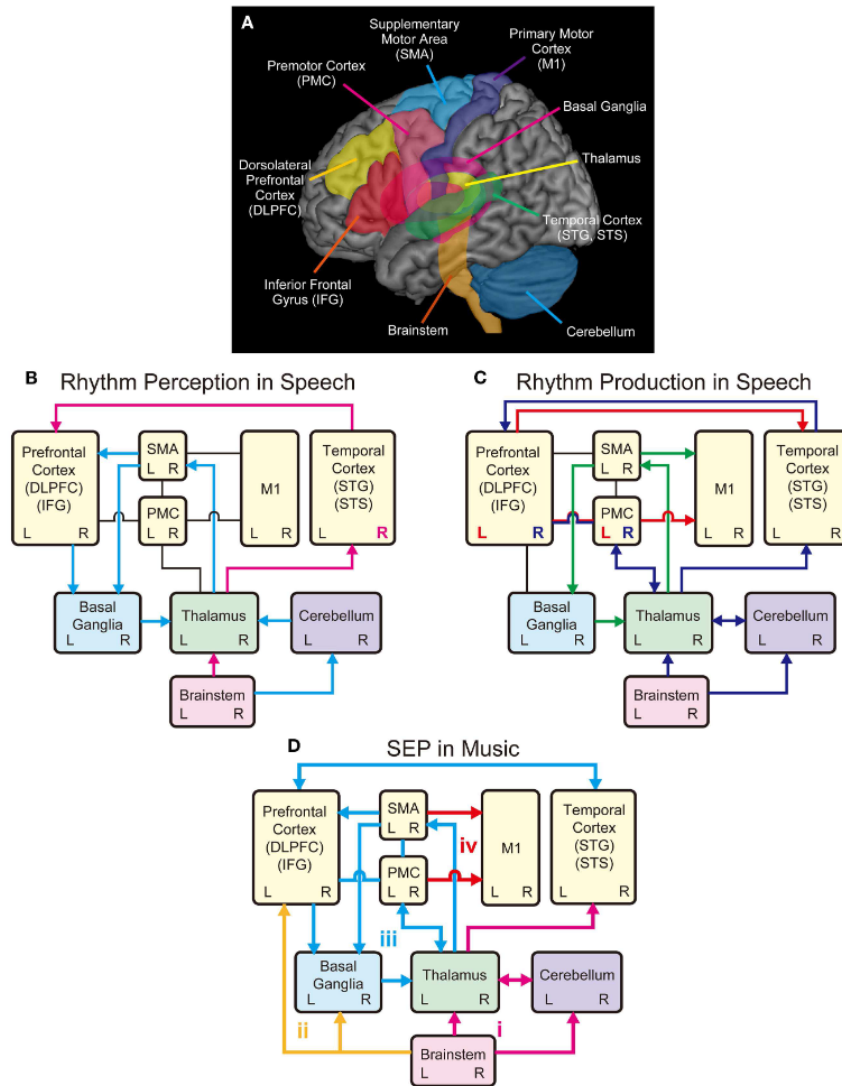
The term MEP was later replaced by "mirror neurons" (Rizzolatti & Arbib, 1998), and is not a metaphor, but an observable physical reality. The patterns of activity are similar in humans, but not identical, and neuroscientists have not yet been able to identify mirror neurons in the human brain. There is however evidence for "mirror networks" or what we can refer to as neural resonance (Hilton, 2022: 34), for example action word processing activating the premotor and motor systems in the brain (Pulvermüller, 2005). We process predictable visual activity unconsciously (Wheatley & Wegner, 2001), and facial expressions are processed particularly quickly (Schindler & Bublatzky, 2020). There is also evidence that observing gestures produces neural entrainment, stimulating activity in the motor areas of the brain (Ping et al., 2014). This idea is at the heart of motor theories of language perception, as we shall see in section 3.5.1. Imitation is a natural instinct in many animals, including humans: we tend to imitate facial expressions and gesture (Rizzolatti & Craighero, 2004) and this is an important part of human social interaction. This mimicry, or "social alignment" (Hilton, 2022: 34) has contributed to the development of human language, and plays an important part in language development in infants, as we shall see later on in this chapter.

### *2.2.2. Prosody and the brain*

The functional lateralization hypothesis (FLH) (van Lancker, 1980: 243) states that processing of prosody depends on its function and the size of the linguistic unit concerned: the processing of prosody relating to highly structuring prosodic forms takes place in the left hemisphere, and the processing of less structured forms, for example those relating to attitude and emotion, takes place in the right hemisphere. Indeed, by analysing dysprosody (i.e. problems processing prosody) in patients with brain lesions, Häuser and Domahs (2014) found that the representation of lexical stress crucially relies on the functioning of the "language-dominant" (mostly left) hemisphere. Imaging studies have shown that



language and music are processed in separate areas of the brain, especially the cortical regions in and around Broca's area and in its right-hemisphere homologue, but that processing language and music involve similar processes (Fedorenko et al., 2009: 2). More recently, fMRI studies have found some evidence for similar lower-order processing, but not in higher-order cognitive mechanisms: Rogalsky and her colleagues (2011) concluded "Music and speech stimuli activated largely distinct neural networks except for in an around core auditory regions, and even in these overlapping regions, distinguishable patterns of activation were found" (Rogalsky et al., 2011: 3849). Recent advances in neuroimaging have also enabled us to visualise the neural pathways along which prosodic information is carried, and it appears that the brain does indeed process prosody and music in a similar fashion. For example, Sammler and colleagues (2015), by using audio morphing combined with multimodal neuroimaging and brain stimulation, found that "prosody perception takes dual routes along dorsal and ventral pathways in the right hemisphere" (Sammler et al., 2015: 3079), and the right side of the brain is also associated with music. Fuji and Wan (2014) examine the data from fMRIs and magnetoencephalography (MEG) and propose the following models for rhythmic perception and rhythmic production in speech (Figure 2.2.2). They demonstrate that rhythmic training can aid stroke patients (Fuji & Wan, 2014). Drawing on research using neuroimaging to measure processes involved in rhythmic speech perception, they identify the neural cues of speech perception and the neural cues of speech production, and propose the SEP hypothesis, which (1) "sound envelope processing" and (2) "synchronization and entrainment to pulse" may help stimulate brain networks that underlie human communication



**Figure 2.2.2. Schematic models of shared brain network for rhythm perception and production in speech and music (Fuji & Wan, 2014: 777)**

In figure 2.2.2, (A) represents possible shared brain regions for rhythm processing in music and speech, (B) is a model for rhythm perception in speech, (C) is a model for rhythm production in speech, and (D) represents a model for Sound Envelope Processing (SEP) and Synchronization and Entrainment to a Pulse (SEP) in music.

The sound envelope processing model provides a framework for understanding how musical training can use neural entrainment to reconnect the neural routes between the prefrontal cortex and the temporal cortex, stimulating both and re-establishing the neural pathways necessary for speech rehabilitation in stroke patients. Language learners do not, of course, have the same problems as stroke patients, but the applications of this model are as relevant to pronunciation instruction for language learners as they are for therapy for stroke patients. As we will see in chapter seven, the

parallels between prosody and music provide many opportunities to incorporate music and certain aspects of the methodology of musicology into research and teaching.

### 2.3. Prosody and age

Age has often been related to pronunciation in L2 learning studies, but before we deal with foreign, second or other languages, in this section, I will focus primarily on the role of prosody in L1 acquisition at various stages in one's life, from before birth, through to old age.

#### 2.3.1. *Prosody in the womb and early infancy*

The human foetus begins to develop senses very early in the womb, the first being touch, at 7 weeks, in the area of the lips and face (Lecanuet & Schaal, 2002: 33), so some vibrations are already being felt at this age. Auditory perception begins when the cochlea begins to develop at 18-20 weeks, and auditory functions are probably complete during the eighth month, when the synaptic connections are organised (Lecanuet & Schaal, 2002: 39). Of course, what the foetus hears will not be exactly the same as what we hear as children or as adults, though air when other people speak to us, or through our own jawbones and skulls when we hear our own speech. When the child's mother speaks, the amniotic fluid, placenta and uterus attenuate the sound, and when someone else speaks, there is also, skin, muscle and fat tissues all of which contribute to attenuating the speech signal and act as a low-pass filter. The exact frequencies of which sounds pass to the foetus' ear can be measured using a hydrophone, which measures decibels and sound pressure levels at various frequencies. The sound pressure levels of long wavelengths and low frequency sounds (<300 Hz) are generally similar in and ex utero, but higher frequency sounds are muffled and significantly attenuated. If an adult listens to recordings of the sounds as the foetus would perceive them, they would hear that the prosodic characteristics are maintained, and that some phonemes can be identified in adults' speech even far from the placenta (Lecanuet & Schaal, 2002: 37). The phonemes in question would mainly be vowels, nasals, etc., with the higher frequency sounds such as plosives and fricatives would be indistinguishable. In effect, therefore, for between two and three months before birth, the human foetus is receiving the intonation and rhythmic patterns of its L1.

The prosodic cues which the foetus is able to hear and store, enable what has been called "phonological bootstrapping" in the first two years of L1 acquisition (Gleitman & Wanner, 1982), and since prosody plays a major role in this process, many authors prefer to refer to "prosodic bootstrapping" (Pinker, 1984; Morgan & Demuth, 1996). The prosodic bootstrapping hypothesis states

that the processing of prosodic information facilitates the acquisition of some structural properties of the language. There have been a number of studies which have explored various aspects of this hypothesis, for example Nazzi, Bertoncini and Mehler (1998) found that rhythmic properties might play a crucial role in new-borns' ability to discriminate between languages. They also found that neonates could discriminate phonetically varied words differing in intonation contour, which suggests that they have already extracted and stored prosodic information (Nazzi et al., 1998: 783).

Judit Gervain and Janet Werker (2021) review studies of prosodic bootstrapping and conclude that prosodic information learned in utero helps infants learn about basic word order, and also aids syntactic analysis and therefore word learning. Recent research using machine learning also provides strong evidence that spectral envelope (phoneme formant structure), pitch features, and rhythm are important for human cognition in infants (Lau et al., 2022). Using samples of female vocalizations typical of speech directed to four-month-old infants, they ran four different models to classify the samples from the vantage point of a four-year-old infant. All four models (spectral envelope, pitch features, rhythm and a combined model) produced statistically significant results. Neuroimaging techniques have enabled the visualisation of brain activity during dialogues between infants and caretakers, and neurons in the auditory cortex oscillate at frequencies that entrain speech rhythm. This entrainment, which enables infants to extract hierarchical information, including lexical stress, syllabic structure and syntactic patterns (Goswami, 2019), evident already at 4 months, develops throughout infancy (Attaheri et al., 2022), and continues to support language processing in adulthood (Poeppel and Assaneo, 2020).

When an infant and a parent or caretaker communicate, there is a form of "rhythmic attunement" (Tomasello et al., 2005), and both interlocutors adjust their rhythm to the speech of the other, thereby continuing the process of prosodic acquisition. Receptive prosody skills are acquired by the end of an infant's first year (Kuhl, 2004: 836). The production of prosody is the subject of much concerted practice by the child during first months of life, as the coordination of articulatory apparatus is achieved through the various stages from non-speech sounds, through babbling, to speech sounds, with the complete phonological system including all of the phonemes usually mastered at around eight years (Kuhl, 2010: 716). The following diagram (Goswami, 2022) borrows a timeline from Patricia Kuhl (2004), and adds the important moments in the L1 acquisition process where key stages in the development of rhythmic acquisition occur:

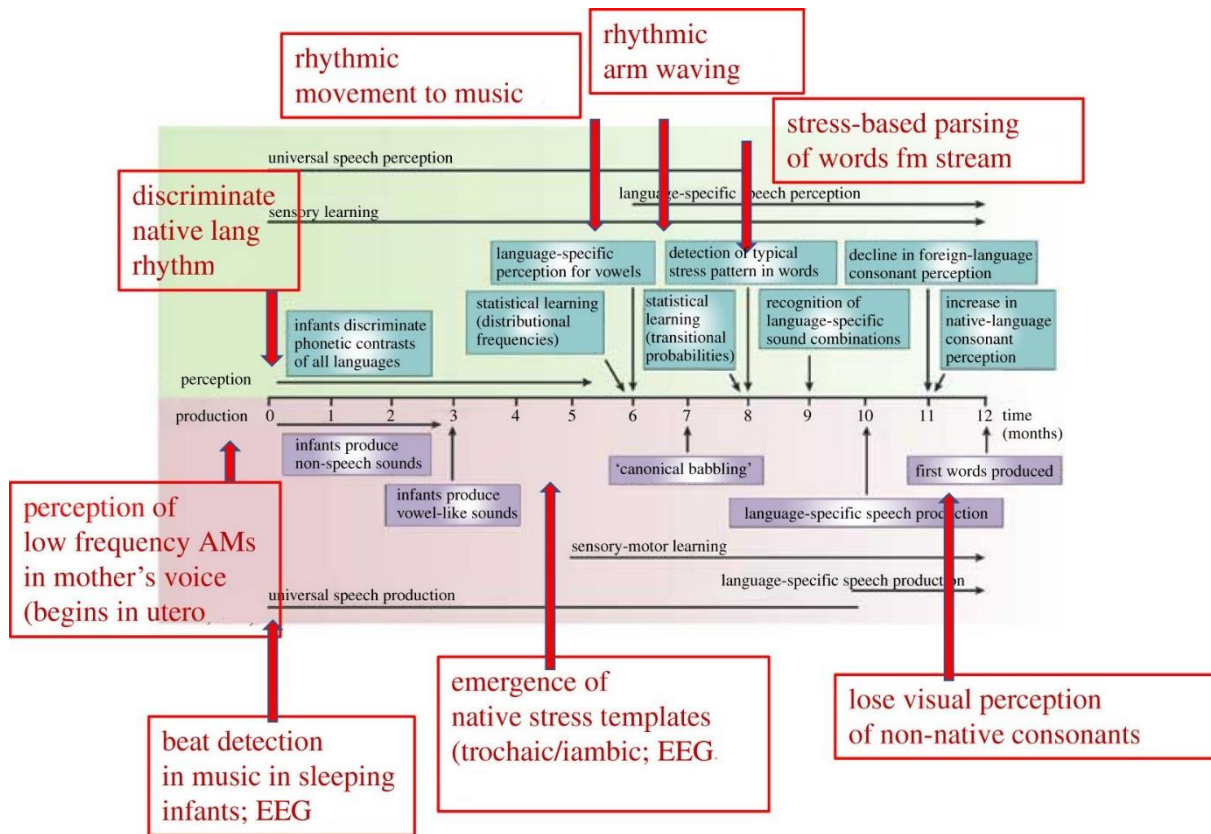


Figure 2.3.1a. The timeline of infant speech acquisition (from Kuhl (2004), with perceptual and motor studies on rhythmic acquisition added (Goswami, 2022: 3))

As the infant goes through the stages from pre-lexical representation of language to lexical access and production, prosody is therefore an essential aid. Anne Christophe and her colleagues (2008) propose a model for representing prosodic bootstrapping, illustrated in figure 2.3.1b:

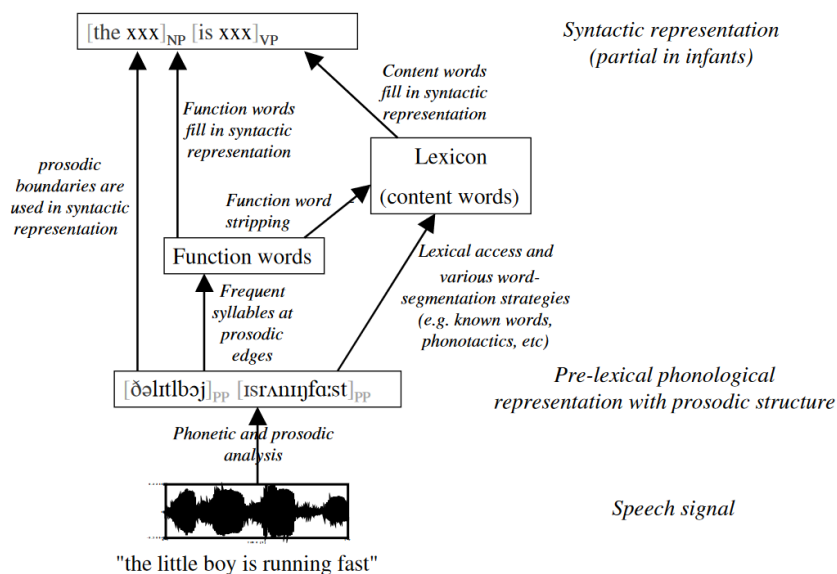


Figure 2.3.1b. A model of speech processing and early language acquisition (Christophe et al., 2008: 62)

This model gives prominence to prosodic information in the pre-lexical stage, much of which, as we have seen, is already stored at birth, and also to the phonetic content of the utterance.

Prosodic bootstrapping has very important implications for teaching and learning LX English, especially for those learners whose prosodic system is very different from English. We begin to perceive the prosody of our L1 through the vibrations of the rhythm of our mother's speech, then from the low frequencies of the speech of other speakers, before we are even born, so that the prosody of an individual's own L1 is very deeply engrained in their neural networks for language. In order to help learners to acquire new prosodic patterns, prosody must therefore be given considerable attention in any teaching programme. Without wishing to suggest that we should try to replicate L1 acquisition conditions for LX learning, anything which can help learners to override hugely engrained perception routines - and help them *feel* the rhythms of the target language - could reproduce, at least to some degree, the first sensations they had of the prosody of their own L1.

Prosody is also an important feature of infant-directed speech. Infant-directed speech has a very specific prosody, essentially exaggerated, to aid acquisition, and this is true across languages (Fernald et al., 1989). Compared to adult-directed speech, infant-directed speech has a slower rate, a higher fundamental frequency, greater variability in amplitude, more lengthening and more pauses (Ma et al., 2011). This enhanced prosody is supposed primarily to enhance lexical and syntactic acquisition: indeed, new words are often introduced in utterance-final position, where they tend to be louder, longer, and higher-pitched (Fernald & Mazzie, 1991). Studies have shown that thanks to prosodic cues, infants as young as six months exhibit rudimentary speech segmentation capabilities (Johnson et al., 2014). Gesture and facial expression are also an important part of infant-directed speech. Combined with prosodic cues, they are also extremely important for language development (Hübscher & Prieto, 2019): in terms of both perception and production, for lexical, syntactic and pragmatic acquisition. Gestures such as pointing, beat gestures, head nods, head shakes, shoulder shrugs and many other gestures have been observed in children, as in adults, corresponding with prosodic prominence and temporal features (Hübscher & Prieto, 2019), and these gestures also aid LX pronunciation learning, as we shall see.

Finally, in the context of L1 acquisition and learning, I would like to underline the importance of certain types of texts for children, notably poems, nursery rhymes, and skipping rhymes. In the same way that infant-directed speech has an exaggerated prosody, and often accompanied by gesture and exaggerated facial expression, so are children's rhymes. We could refer to these prosodically rich and culturally important formalised as examples of prototypical speech (Cauldwell, 1996; Lin, 2012), most notably in their rhythmical structure. In focusing children's attention on the prosody of their L1 though

bedtime stories, play, etc., we aid the acquisition of the rhythmical structure of the language, which will help them with processing oral language and to speak more fluently. Children's rhymes are of course present in all languages, and there is a particularly long and rich tradition of poetry for children in the English language, from Edward Lear, A.A. Milne and Lewis Carroll, to Dr. Seuss, Spike Milligan and Julia Donaldson, all with a very regular rhythmic component. The rhythm of English poetry has its origins in some of our earliest fiction, the scopos of old English and Norse, and early poems of other Germanic languages, from which we have inherited so much of our rhythmic structure. As Lehmann (1956) explains:

The four-beat line has great historical depth and appears to be linked to the earliest poetry in the Germanic languages, in which the line is made up of four predominant syllables, "[...]two in each halfline, which are elevated by stress, quantity, and two or three of them by alliteration. (Lehmann 1956: 37)

The rhymes which we read our children and which children use themselves in play are some of the most rhythmically regular styles of speech which exist in English. Andy Arleo (1997) developed a hypothesis of metrical symmetry (based on a large corpus of British and American children's rhymes). He revised his hypothesis to the following:

Children's rhymes tend toward symmetry, defined as follows:

1a. Beats (version a). The number of beats in a given metrical unit (i.e., hemistich, line, stanza) tends to be even.

1b. Beats (version b). The number of beats in a given metrical unit tends to be a power of two ( $2^n$ , where  $n > 0$ )

2a. Lines (version a). The number of lines in stanzas tends to be even.

2b. Lines (version b). The number of lines in stanzas tends to be a power of two. (Arleo, 2006: 48)

Arleo also studied a corpus of French and English "counting out rhymes" (the rhymes children use to choose who will or will not participate in a game, for example) and found that they are rhythmically different. In English, the number of lines is generally even and equal to a power of two, and tend to have four beats per line, however he states that "Evidence from French counting-out rhymes is not as clear, although there is a slight preference for stanzas with an even number of lines and for lines with an even number of beats." (Arleo, 2006: 16). We may question why this is the case, and in her analysis of a corpus of 130 English nursery rhymes, Jacqueline Guéron (1974) offers a possible explanation. She concluded that the metrical patterns she found could "only be explained by the similarity between the metrical rules for the nursery rhyme line and the rules for English clausal stress" (Guéron, 1974: 111). For these reasons, as I will outline in chapter seven, the use of nursery rhymes in teaching English as an L2, especially in France, can be particularly useful. Nursery rhymes are generally associated with younger learners, but they can be a useful aid for learning new prosodic patterns all through a learner's life.

### *2.3.2. Prosody, accent and aging*

Our accents are something which evolve over our lifetimes, sometimes consciously, sometimes unconsciously. Linguists are increasingly realizing that the ways in which the components of children's language interact, are crucially important to analytic characterizations of adult language (Shatz, 2007: 4). However, it is when we are infants that we acquire our accents which, to a greater or lesser degree, stay with us throughout our lives. The critical period hypothesis (Lenneberg, 1967), also known as the sensitive period hypothesis also states that languages are more easily learned in childhood before puberty sets in, particularly regarding pronunciation. This is a phenomenon which has been explored more deeply in L2 learning and acquisition than in L1 acquisition, and we will look at some of the issues around the critical period hypothesis in the next chapter.

As we get much older and our cognitive processes deteriorate, studies have shown that prosody remains important for parsing syntactic structures (Wingfield et al., 1992), although some research has shown that older subjects are less likely to pick up on affect expressed via prosodic cues, and tend to use semantics and syntactic cues more (Ben-David et al., 2019). This being said, Boaz Ben-David and his colleagues also note that this may be because of hearing degradation, and that in addition, it is difficult to separate the two channels of prosodic and semantic cues. Regarding production in older language users, problems are more likely to be found in the area of lexical access, such as remembering words (Burke & Shafto, 2004). Of course, the natural process of aging results in a deterioration of muscles and other tissues responsible for phonation, leading to a perceived hoarseness, breathiness, reduction in F0 in women, etc. (Karlsson & Hartelius, 2021), but other than speech rate and pausing, prosodic patterns remain relatively unchanged. In fact, even in dementias, such as Alzheimer's, prosody is one of the last features of speech to disappear, although a flattening of the F0 curve, more frequent and longer pauses, and increased speech rate between pauses (possibly as a compensatory technique) may be indicators of the onset of dementia (Farrús et al, 2020).

### *2.3.3. Prosody in bilingual infants*

As I mentioned in the introduction, although many of us in the Western World may think monolingualism is normal, over 50% of the world's population speaks more than one language or dialect in their everyday life (Grosjean, 2021), and many learners of L2 English have already acquired several languages before formal schooling begins. With increased geographical mobility, an increasing number of students are already bilingual when they reach universities in France, be they French or



international students, and therefore the question of bilingualism is relevant. For bilingual children, the picture is complicated when it comes to phonological acquisition, and a certain amount of interaction between languages during the acquisition process is inevitable. There may be social reasons for wanting to raise one's child as a monolingual, but a number of false assumptions are still held by parents, teachers, and even doctors about raising children as bilinguals. However, in an extensive review of studies relating to bilingual infants, Fred Genesee (2008) concluded that there was sufficient evidence to dispel four commonly held myths about language acquisition and language learning, namely the monolingual brain; that younger is better; time-on-task; and that bilingualism is not advisable for children with developmental disorders or academic challenges.

Johanne Paradis (2007: 388-89) highlights the fact that children's phonological acquisition is faster and more effective than that of adults who become bilingual later in life. Although in a later study, she focuses on lexical and morphological acquisition in bilingual children (2011), and concludes that a usage-based theory would be a more likely explanation than a biological theory. There is evidence that bilingual infants, aged 7 months, are able to distinguish word order in both of their languages based on prosodic cues, and so prosodic bootstrapping appears to aid them in linguistic processing and acquisition of both languages, as it would monolingual children (Gervain & Werker, 2013). As a language teacher, knowing which other languages a learner has spoken since an early age can sometimes help to explain prosodic features in a target language, as linguistic transfer can be made to work in the learner's favour.

#### 2.4. Prosody & sex

Socio-cultural gender differences in language is a vast topic, and I will concentrate on prosody and differences in sex (the difference between sex and gender, and related questions, though important, are not within the scope of this report). In some languages, there are clear differences in the prosody of male and female speech built into the language, as in Asian languages which have separate polite forms for men and women. However, this is not so of English, even though it has long been known that there are some differences in the language of male and female speakers (Trudgill, 1972: 182-3, Lakoff, 1973). First, there are some biological differences between men and women which account for certain differences in male and female speech, most importantly the length of the vocal tract: in men, typically 17-18 cms, and in women, an average of 14 to 14.5 cms (Simpson, 2009). This difference partially accounts for the higher average pitch of women's voices, but when we look at average F0 values, the pitch differences between male and female voices is greater than mere biological differences can account for, so they must be partly learned (Simpson, 2009: 625). Several studies have also shown that

women use more varied intonation patterns and have a greater F0 range in European and non-European languages (Traunmüller and Eriksson, 1995). A possible explanation for a greater F0 range in female speakers is that women have more listener-oriented speech, pay more attention to turn-taking, ask more questions, etc., and therefore use a greater variety of intonation contour patterns, including higher tones (Coates, 2015). There is consensus that any superior verbal performance in females compared to males, including greater observed range in intonation contour patterns, is due to social factors and context, rather than an innate superiority in the female brain (Hyde & Linn, 1988; Sunderland, 1998).

In the context of the French language, a study comparing male and female speakers in American English and French, found a significantly greater range in the female French speakers' F0, whereas there was no significant difference between the male and female speakers of American English.

When it comes to studying languages, females choose languages more often than males. Girls choose to study languages more often than boys at school (Sunderland, 2000). Furthermore, a 2020 British Council report also shows that not only are girls more likely to study languages at school in the UK, they also outperform boys considerably (Collen, 2020). This carries through to university: The Higher Education Statistics Agency (HESA) figures show that around 70% of enrolments on language courses in the UK have been female over the last three years, and similar figures can be seen in French universities.

Regarding prosody, sex and L2 learning, there is scope for more research. Gabriel and Kireva (2014), found no measurable differences in the speech of male and female speakers with respect to the realization of prenuclear accents and final contours (Buenos Aires Spanish and Castilian Spanish produced by Italian L1 speakers). Finally, in my own research carried out for my PhD, sex was one of the independent variables in my study. Although I found no significant effect on learning outcomes, I did find the goal-oriented motivation to improve pronunciation to be statistically greater for the female participants in the study (18-20-year-old IT students).

## 2.5. Prosody, accent and identity

By accent, I refer here to L1 accents, not L2 accents, which I will deal with in the next chapter and in chapter six in more detail. The way we speak - the language we use, how "correct" our grammar, how strong our dialect, etc. - is a very strong marker of identity, and as Hymes (1974) points out, accent is an important part of this:

Any enduring social relationship or group may come to define itself by selection and/or creation of linguistic features, and a difference of accent may be as important at one boundary as a difference of grammar at another. Part of the creativity of users of languages lies in the freedom to determine what and how much linguistic difference matters. (Hymes, 1974, 123)

The way we speak in our L1 may provide information about where we come from, our class, our education, and much more (Setter, 2019). Prosody is, of course, one of the features of accent that may be perceived by others as a marker of identity. Whilst the role of stress and rhythm is perhaps less distinctive in this regard, intonation is often identified as a social marker, as Wells (2003: 12), and the use of HRT in English is frequently discussed example of this. Our accents, and the prosodic features which are an intrinsic part of them, contribute to the judgements people make about us when we are speaking our own languages, and this has implications for our personal relationships and our professional lives. For these reasons, it is hardly surprising that many people modify their accents over time in order to better fit into the community in which they find themselves, including English LX teachers (Baratta, 2018).

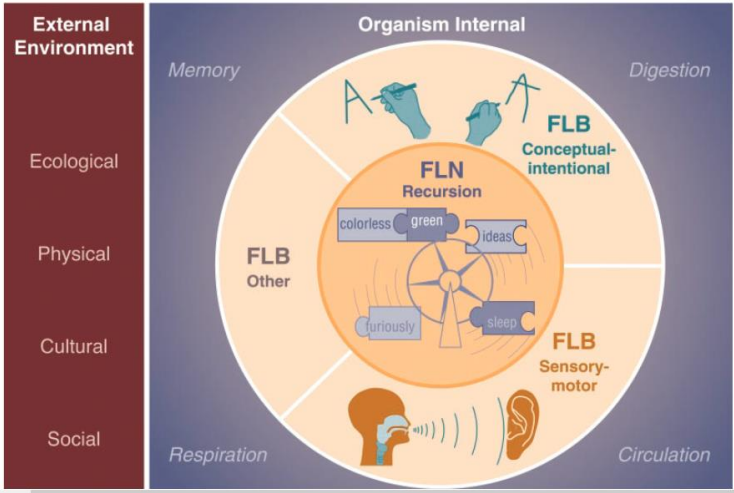
Prosody also plays a key role for the perceived persuasiveness, self-confidence, and passion – in short, the charisma – of a public speaker (Wörtwein et al., 2015). While it is perhaps impossible to teach charisma, the prosodic aspects which contribute to it may certainly be learned; Oliver Niebuhr and his colleagues have proven that with only an hour of training, statistically significant results may be attained in the six prosodic features they consider essential to charismatic speaking (Niebuhr, 2021).

## 2.6. Prosody and the origin of human language

In this final section of the chapter, we shall explore the development not of language within an individual's lifetime, but the origins of human language itself, in order to illustrate the important role of the vocal and physical cues to prosody in this process. By examining research from different fields, we will see how the human body, brain and language have evolved together, and that prosody has been at the centre of this process. Understanding some of the processes which have developed over millions of years (such as social alignment, rhythmic entrainment, and framing) can aid teaching, and help learners to think about the nature and roles of prosody.

As the speech apparatus does not fossilize, we can only make inferences as to how language developed from fossil evidence and from the tools and other artefacts we have found from prehistoric humans from tens of thousands to millions of years ago. From evidence from the fossil records (Mithen, 2011), along with what we can see in human behaviour today, and in the behaviour of our closest living relatives, apes and monkeys, it is possible to make hypotheses about the origins of human language. As Hauser, Chomsky and Fitch (2002) point out, hypotheses as to the origins of human language “can

be distinguished only by empirical data, much of which is currently unavailable” (Hauser et al., 2002: 1573). They call for a collaborative and comparative approach, drawing on the relevant branches of linguistics, biology, psychology, and anthropology to construct testable hypotheses (Hauser et al., 2002: 1578).



**Figure 2.6. A schematic representation of organism-external and -internal factors related to the faculty of language (Hauser et al., 2002: 1570)**

The model represented in figure 2.6 represents some of the factors which the authors suggest can be explored in researching the development of the faculty of language in its broad sense (FLB), and the core grammatical computations included in the development of the faculty of language in its narrow sense (FLN). The model includes sensory-motor, conceptual-intentional, and other possible systems (which the authors leave open) involved in the development of language. In this section, we will examine some hypotheses from different fields concerning the “organism internal” and “organism external” factors which have contributed to the evolution of human language, and to our understanding of the nature and roles of prosody.

*2.6.1. The musical protolanguage and gestural protolanguage hypotheses*

One argument which has been advanced for prosody being an early feature of human language is its presence in non-human primate communication (MacMahon, 2007). Chimpanzees separated from our human ancestors some 8 million years ago, and *Australopithecus* evolved some 4 million years ago, spreading throughout Africa and beyond, but *Homo habilis*, our first direct ancestor, evolved around 2.8 million years ago (Fitch, 2010, chapter six). Their brains were about the size of a chimpanzee’s brain, but they used stone tools and we assume that their language was more developed than chimpanzees, who use vocalizations, facial gestures and physical gestures to communicate. Apes,

monkeys and animals other than humans certainly have language, and there are many documented cases of apes and other animals using human sign language and other forms of human communication (Fitch, 2010: 166), but human language is more complex than any other language that we know of. More precisely, our speech apparatus and specifically our vocal tracts, allow for more complex formant structures than in other animals (Fitch, 2010: 310). In addition to the study of our ancestors through fossil remains, we can also examine language in living primates, and make inferences about how our own gestures, vocalizations, and language processing developed (Vauclair and Meguerditchian, 2008). Rizzolatti & Arbib, (1998), cited in the previous section for their work on mirror neurons, argue that the work they carried out on monkeys and humans supports the hypothesis that facial expressions and gestures were at the origins of human language:

(1) the mimetic capacity inherent to F5 and Broca's area had the potential to produce various types of closed systems related to the different types of motor fields present in that area (hand, mouth and larynx); (2) the first open system to evolve en route to human speech was a manual gestural system that exploited the observation and execution matching system described earlier; and (3) that this paved the way for the evolution of the open vocalization system we know as speech. (Rizzolatti & Arbib, 1998: 192)

It is the opinion of Rizzolatti and his colleagues that manual gestures gradually lost their importance as vocalisation became more prevalent and a more effective way of communicating needs to others (Rizzolatti & Arbib, 1998: 193). This point of view, that gesture preceded vocalization at the origins of language is typical of one school of thought, whereas another school of thought hypothesizes that vocalisations precede gesture. Darwin suggests in *The Descent of Man* that human language developed from a protolanguage which shared many characteristics with singing mentions "musical notes and rhythm" and explains that "When we treat of sexual selection we shall see that primeval man, or rather some early progenitor of man, probably used his voice largely, as does one of the gibbon-apes at the present day, in producing true musical cadences, that is in singing" (Darwin, 1891: 56). Darwin's "musical prototype" hypothesis (Fitch, 2013: 492) does not however exclude early use of gestures. As Darwin also states: "I cannot doubt that language owes its origin to the imitation and modification, aided by signs and gestures, of various natural sounds, the voices of other animals, and man's own instinctive cries" (Darwin, 1891: 56). Since Darwin's writings on the subject of the genesis of human language, many authors on the subject fall into one of two camps: there are still those who hold that human protolanguage began with gesture (e.g. Hewes, 1973), and proponents of a musical theory of human protolanguage (e.g., Mithen, 2005). We cannot of course be sure whether human protolanguage began with gesture and facial expressions, or whether it began with vocalizations, indeed it is more likely that it was a combination of the two, with instinctive vocalizations and physical movements evolving into the complex behaviour and "neuromechanics" of human language (Chung, 2018). So, before our speech organs developed sufficiently to start producing the complex formant patterns which make human language so infinitely varied, there were gesture, facial expression and

variations in the frequency, amplitude and duration of the sounds we uttered, and that these corresponded to meaning, to new and important information our ancestors wished to convey to others: prosody is at the very beginnings of human language.

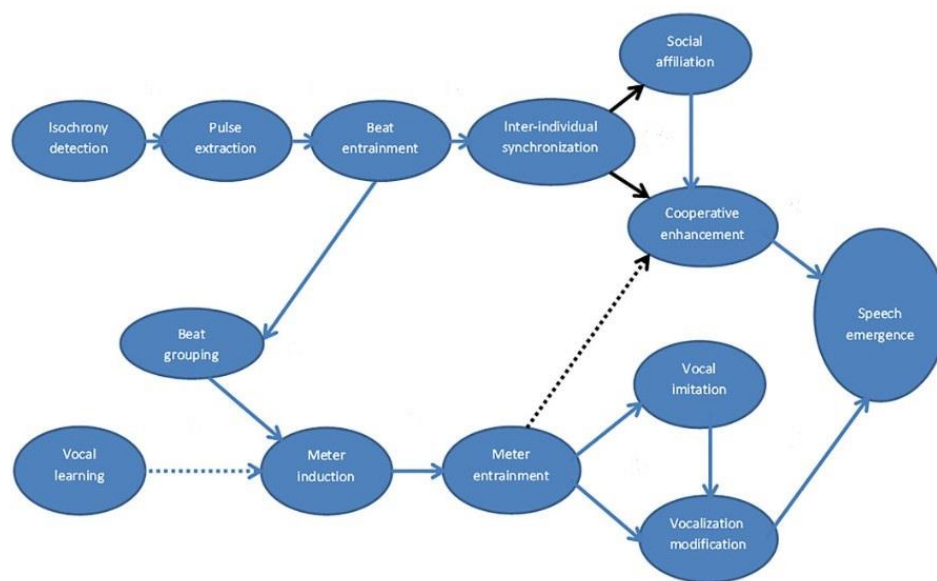
### *2.6.2. Bodily rhythms and the rhythms of language*

Given then, the probable links between gesture and vocalization at the beginnings of human language, and which are still present today in prosodic cues, several authors have hypothesised that the rhythms of the body have influenced the rhythms of human speech. Peter MacNeilage's frame/content theory (Davis & Macneilage, 1995; Macneilage, 1998) describes a possible causal relationship between the biphasal rhythms of our body – such as the opening and closing of the mouth – which are necessary to our producing sound, and the origins of the rhythm of human speech. Another physical constraint of speech is breathing. Lung capacity imposes limits on the length of spoken sentences (Hauser et al., 2002: 1571), and the time needed to breathe in separates our speech into breath groups, or tone units. An adult breathes at around 12 to 20 times per minute, depending on activity (Barbosa Pereira et al., 2017), representing three to five seconds per cycle. This corresponds to the average duration of breath groups, which Yu-Tsai Wang and colleagues measured at 3.5 seconds for reading aloud and 4.35 seconds for spontaneous speech (Wang et al., 2010).

Other rhythms may have influenced the development of human speech, such as the walking and heartbeat. The preferred walking speed for humans ranges from 75 to 125 steps per minute (Sabatier & Ekimov, 2008), and in laboratory studies this varies, but the preferred cadence has been shown to be around 120 steps per minute (MacDougall & Moore, 2005). The adult heart beats within a similar range: at rest it beats at around 75 beats per minute, and rises to 100-120 beats per minute with moderate exercise (Tortora & Derrickson, 2009: 741). It is an innate quality of the human mind to impose a perceived rhythm on any periodic event, as a series of experiments in the 1950s and 1960s showed (Allen, 1975: 76). When we walk in pairs or in groups, we often unconsciously align our gait with our partner when walking together (Larsson, 109: 4), an example of the phenomenon of social alignment, which we saw in section 2.2. It is plausible that the rhythm of human language developed along with these other bodily rhythms, as several authors have hypothesised, through attested practices such as “pulse extraction” and “beat entrainment” (Fitch, 2011). Polyanskaya and her colleagues (2019) devised a series of experiments to investigate the relationship between cooperation and rhythm perception to see if “rhythmic synchronization” operates in groups engaged in cooperative behaviour. Their study confirmed links between social interactions and speech patterns, and they draw the following conclusions from their study:

The mapping of vocal rhythm convergence onto social affiliation is important for the development of social cognition and for language acquisition in ontogenesis and probably was an important facilitating factor for speech emergence in phylogenesis. (Polyanskaya et al., 2019: 8).

In figure 2.6.2., Leona Polyanskaya and her colleagues (2019: 8) represents the rhythmic processes leading to the emergence of speech. In figure 2.6.2, we see the parallel processes of group learning (along the top of the diagram) and individual learning (along the bottom). Dotted lines stand for controversial causal links (i.e., those for which no empirical evidence or inconsistent empirical evidence exists), solid lines stand for experimentally tested hypotheses and the black lines represent the links in the paper in which this diagram appears:



**Figure 2.6.2. The role of rhythmic cognition in speech emergence from an evolutionary perspective. (Polyanskaya et al., 2019: 8)**

These two processes are of course different, but in the context the development of prosody, the role of rhythm is key in both. As Tomassi and his colleagues remind us, matched social dialogue of human infants and their mothers (or caretakers) is uniquely human – no other species combines facial social signals and meaning in this sort of “rhythmic face-to-face exchange” to this extent (Tomasello et al., 2005: 698). These extremely prosodically rich exchanges between infants and mothers, along with the other forms of rhythmic social alignment of individuals as they grew into adults and walked, talked, sang and made music together have all contributed to the development of human speech. We will present the concept of rhythmical entrainment in more detail in chapter seven, along with some of its pedagogical implications.

### 2.6.3. *The prosody of Proto-Indo-European*

We know from the analysis of fossil evidence and the analysis of mitochondrial DNA samples, that our earliest ancestors began their expansion from Africa 59,000 to 69,000 years ago (Maca-Meyer et al., 2001). Two main groups formed in Africa, one that spread to North and east Asia and India some 30,000–57,600 years ago, and one 43,000–53,000 years ago, that spread from western Asia, bringing Caucasians into North Africa and Europe. After a period of co-existence and some interbreeding with Neanderthals, humans arrived in Northern Europe, as far as the Arctic Circle, around 40,000 years ago, and that the colonization of this region by modern humans happened 13,000-14,000 years ago (Pavlov et al., 2001; Higham et al., 2011). For tens of thousands of years, therefore, these early humans led a semi-nomadic existence, walking, talking, nursing their babies, telling stories, singing, playing music and dancing. The last places they reached, the furthest from their origins in Africa, were the frozen tundra and the vast forests of Northern Europe, where gradually their languages evolved into the earliest language that linguists have been able to reconstruct by working back from languages which we have some record of: Proto-Indo-European. Although we have no direct record of Proto-Indo-European, it is thought that there was a strong/weak alternation, i.e. it was a language which marked stress (Lehmann, 1952). The prosody of the earliest “known” European language, albeit reconstructed, through a gradual process of rhythmical entrainment, was shaped by the combined rhythms of heartbeat, walking and breathing, in much the same way that each of us first experiences rhythm in the womb.

In this chapter, we have seen that certain aspects of prosody, notably those relating to the cue of F0, are governed at least partially by biology, but that biology alone cannot explain the way we use prosody both in terms of perception and of production. When we learn our first language or languages, we start by acquiring prosodic features several months before being born, and continuing through early infancy, primarily concerns the acquisition of basic prosodic structures, which this helps us to acquire basic syntax and aids lexical acquisition in infancy. These auditory and physical cues to prosody, which have developed over millions of years as humans have interacted with their environment, are some of the last features of language to persist into an individual’s old age, and even into the advanced stages of dementia. When a new language is learned, all of the factors we have seen in this chapter potentially influence that learning process, and in the next chapter, we shall examine the role of prosody in LX learning.



## CHAPTER THREE

### LX prosody

In this chapter, we will explore some of the issues which arise when languages come into contact through language learning. I will leave questions of learning outcomes and pedagogical choices until part two of this report, focusing instead on factors which influence the *process* of learning. While continuing to focus on prosody, we will firstly examine how certain individual factors and societal factors influence LX learning. Then, after exploring some key questions in language transfer, we will look at some models of language learning in which phonological transfer plays an important role. Questions pertaining to the particular differences between English and French will be dealt with in the following chapter.

#### 3.1. Social factors in LX learning

As we saw in chapter two, accent is a powerful marker of identity, for language users as individuals, and also as members of a wider society. This is as true for LX users as it is for people in the language or languages they acquired from birth or from an early age. As Nunan (2013) points out, these psychological and societal factors are ever-changing and infinitely complex, and attitudes towards learning language are unstable and constantly evolving. Within this context, the role of accent, and the processes and the goals of language learning are intimately interconnected with other aspects of learners' lives. Secondly, learner difference is a complex construct that cannot be simply reduced to the influence of isolated variables.

#### 3.2. Individual differences

It is now widely understood that learner-centered teaching - defined as when teachers "know their learners well and [are] responsive to their needs and preferences in language learning" (Benson, 2012: 30) - is beneficial to learning outcomes. However, regarding accent, this was not always the case, as we see in section 7.1. Even today in many contexts, particularly in universities where large class sizes and other institutional constraints make learner-centered teaching difficult, learner-centered teaching is not always practised. Education research may also benefit from a learner-centered approach, and yet many studies choose to focus on quantitative data and controlled studies, where individual learner

differences are often glossed over. However, in a mixed methods approach, it is possible to focus on individual learner issues by complementing quantitative research with qualitative data-gathering, such as questionnaires and guided interviews. In this way, research can be more learner-centered. An example of this is the ELLO project I ran with my colleague Jean O'Donnell in Savoie Mont Blanc University from 2011 to 2015 (Frost & O'Donnell, 2013; 2015). The study focused primarily on self-assessment and peer-assessment of oral English, however we were also interested in the attitude of learners to their own performances, how they evolved over time, and how this impacted their motivation to improve. The quantitative data we gathered gave us no insight into these questions, and so the participants completed online questionnaires, and a sample of them also completed guided interviews. Mixed methods approaches which explore quantitative and qualitative data have become more popular in language teaching research this century, and better allow researchers to account for individual diversity (Dörnyei, 2007: 27). An embodied cognitive approach in research - as in teaching – means considering learners' thoughts, feelings and interaction with their environments, and this is difficult to achieve simply by measuring learning outcomes and administering learner questionnaires. Quantitative data is essential for comparing learning outcomes in a controlled study for example, and certain cognitive and affective factors may be anticipated and explored with closed questionnaire items, however a truly learner-centered approach to research must leave more opportunities to participants to express themselves.

### 3.2.1. Age

One of the most important individual variables concerning LX phonological acquisition is age. Perhaps the most frequently discussed concept regarding language learning and age, is that of the critical period hypothesis mentioned briefly in the previous chapter (Lenneberg, 1967). Some authors prefer to talk about a "sensitive period" indeed both of these terms have co-existed in developmental biology and developmental psychology for some time (Ruben, 1997). There are critical or sensitive periods for all aspects of physical, cognitive and emotional development, from the earliest years through to the end of adolescence, but as regards language, we can class them into three areas: phonology, semantics and syntax. For phonology, this is considered to be from the sixth month of foetal life to 12 months, up to four years for syntax and through to 15 or 16 years for semantics (Ruben, 1997: 202).

The view that language acquisition may only happen successfully during early life has been challenged ever since the terms of "critical period" and "sensitive period" were first used. Oyama (1976) remarked that the sensitive period is not biological, and that the ability to "speak like a native" is not an "all-or-nothing phenomenon", although she conceded that "seems to be quite difficult for all but the very

young” (Oyama, 1976: 278). Many studies since then have found that learning a second or other languages in the early years of one’s life, or during adolescence, may lead to greater phonological mastery. Patowski (1990) found more LX learners who started before the age of 15 had better accents in the target language, but other authors have refuted this. James Flege (1995), in his influential speech learning model, which I will present briefly at the end of this chapter, considered age to be a key factor in LX learning, however he later revised his model (Flege, 2018; 2021), in the light of more recent evidence including neuroimaging, highlighting the importance of other factors, primarily input. Wendy Baker and her colleagues (2008) conducted two experiments with Korean immigrants learning English earlier and later in life, and found that children outperformed adults in production tests, but that adults performed better at the perception of vowel categories. Age was therefore found to be important in this study, but “age-bound neurobiological limitations” were not as important as prior language learning experiences and language exposure (Baker et al., 2008: 338). Few studies have focused primarily on age and prosodic features in LX learning, but it does appear that age plays a role for both the acquisition of intonation (Ryalls, et al., 1994) and speech rhythm (Szakay, 2006).

More recently, advances in neuroimaging and research into brain plasticity have shown that assumptions that brain plasticity diminishes as the brain ages are simplistic. Greenwood (2007) accounts for the aging process and its effects on cognition with several stages. Initially, there is an age-related loss in the number of synapses and a certain white matter degeneration in adults. Secondly, age-related deficits lead to changed processing strategies. Thirdly, these changed processing strategies in older individuals result in changed cortical innervation (either due to the atrophy or to the changed strategies). And finally, thanks to this “strategy-induced plasticity”, “aged individuals show increased activity in neuroimaging in regions adjacent and contralateral (PFC) to the atrophic area” (Greenwood, 2007: 667). Several studies have shown that brain plasticity training programmes, even for aged subjects, can produce positive results (Mahncke et al. 2006).

And there is plenty of evidence, of course, that pronunciation training programmes for adults, even over short periods of time, can be very effective, as we shall see in the second part of this report. In their concluding remarks to their critical review of recent studies on age-related factors in language learning, Muñoz and Singleton (2011) conclude that although the importance of the age factor is an undeniable reality, “factors other than maturational should be brought more to the fore and treated more seriously” (Muñoz & Singleton, 2011: 24). One of the most important factors in learning to modify or add to one’s pronunciation store is simply having the time and mental space to do the necessary “work” to learn – i.e. exposure to the language, and practice of reception and production skills. This is easier when one is young and has all the time in the world, but not impossible when one is older. If we thought it was impossible to make inroads into fossilized LX pronunciation patterns, then there would

be little point in working in lifelong learning. Fortunately, even “fossilized” features can be improved, (thereby improving intelligibility) as Tracey Derwing and Murray Munro’s (1997) study showed clearly over a 12-week course (Derwing et al, 1997) In fact, some studies, (Bongaerts et al., 2000) have even found that learners may achieve a “nativelike accent in a non-primary language” well after the so-called “critical period” (Bongaerts et al., 2000: 305). In the study, Theo Bongaerts and his colleagues conducted, it was shown that certain immigrants to the Netherlands with an extremely high integrative motivation can lose their L1 accents all but completely. From my experience in research and teaching older learners, I would say that motivation and opportunity are greater drivers of both LX pronunciation learning and more informal acquisition. All learners are different, with different abilities, motivations, and opportunities to improve, and in the next section, we will explore LX learning from the perspective of the learner.

### *3.2.2. Learner profiles and cognitive styles*

As I mentioned in the introduction to this chapter, at the heart of a learner-centered approach to language teaching is getting to know one’s learners, as they are all different. Various attempts have been made to model the different ways in which learners encode information, resulting in a plethora of articles and books on learner styles, learner profiles, cognitive styles, etc. in the last two decades of the twentieth century. Frank Coffield and his colleagues (2004) list 61 models which they categorised into five types: constitutionally-based learning styles and preferences, cognitive structure, stable personality type, ‘flexibly stable’ learning preferences, and learning approaches and strategies (Coffield et al. 2004: 9). Many of these references are to research papers, but more often than not, these theories and models are not founded in empirical research. As Dörnyei (2007) states: “The crux of the problem is that style research in the past has not been able to demonstrate sufficiently that the notion of cognitive style is a “theoretical construct in its own right” (Dörnyei, 2007: 126). Certainly, Coffield and his team found no empirical evidence for any biological foundation for theories of learner styles (Coffield et al., 2004: 12), but they cite many research programmes which give empirical evidence for “pedagogical impact” of application of learning style theories in teaching programmes. The notion of learner profiles, or cognitive styles, has therefore always been both problematic and useful. In the opinion of Ellis (1994), this is partly due to learner cognitive flexibility, and partly to ill-defined research objectives:

At the moment there are few general conclusions that can be drawn from the research on learning style. Learners clearly differ enormously in their preferred approach to L2 learning, but it is impossible to say which learning style works best. Quite possibly it is learners who display flexibility who are most successful, but there is no real evidence yet for such a conclusion. One of the major problems is that the concept of ‘learning style’ is ill-defined, apparently overlapping with other individual differences of both

an affective and a cognitive nature. It is unlikely that much progress will be made until researchers know what it is they want to measure. (Ellis, 1994: 208)

If we turn to cognitive psychology, we see that there have been many attempts to model human personality, or personality types which have been more carefully formulated and attested by experimental studies than those in the field of LX learning. The most influential of these models in cognitive psychology is known as the five-factor model, or “the big five” (John et al., 2008). This model was originated in the 1980s, but is often credited to Donald Fiske (1949), who identified the following five traits: "Social Adaptability", "Emotional Control", "Conformity", "Inquiring Intellect", and "Confident Self-expression." The model has been refined over the following years, but the five initial categories still remain influential today). They are:

- openness to experience (inventive/curious vs. consistent/cautious)
- conscientiousness (efficient/organized vs. extravagant/careless)
- extraversion (outgoing/energetic vs. solitary/reserved)
- agreeableness (friendly/compassionate vs. critical/rational)
- neuroticism (sensitive/nervous vs. resilient/confident) (Roccas et al., 2002)

Research in cognitive psychology relating to personality traits can provide useful insights into language learning, and more research into how personality and learning pronunciation could help stakeholders to make more informed decisions.

### 3.2.3. *Motivation*

Motivation is a very important factor to successful learning in any context, but like many researchers and teachers in the late nineties with an interest in technology-mediated learning, the question of motivation was even more important to me: in autonomous learning, there is no teacher present to help to motivate learners, and so other ways to motivate learners must be found. Sekhar and colleagues (2013), in an extensive literature review on motivation, counted between 1 549 and 10 679 articles or reviews on motivation per year from 1990 to 2013 (Sekhar et al, 2013: 474). My interest in this subject led me to include a questionnaire on motivation in my PhD study, and I also co-organised a conference on the theme of motivation at this time. In 2005, I was tasked with organising the scientific committee for the APLIUT<sup>8</sup> and I suggested the theme of motivation and language learning. As part of the conference, we interviewed Zoltan Dörnyei, and one of his remarks during that interview confirmed my views on the main challenge for developing digital language learning resources – he said

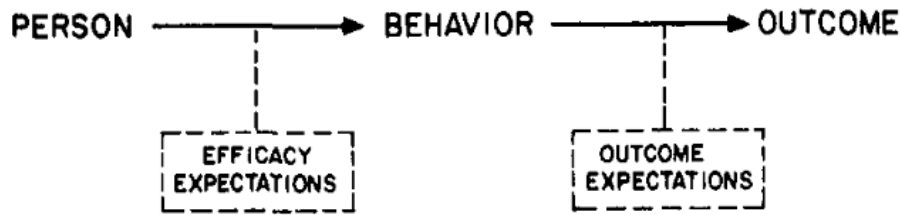
---

<sup>8</sup> *Association des professeurs des langues des instituts universitaires de technologie.*

that that teacher was the single most motivating factor in any language classroom. The other plenary speakers I invited to the conference were Robert Vallerand, whose interests are motivation in education in Canada, (Vallerand, 1997), and Boris Cyrulnik, a psychiatrist, psychologist, neuropsychologist, and ethnologist who has published widely on child and adult psychology. (Cyrulnik, 2022). Reading the work of these researchers, and subsequent reading on motivation, has led me to explore motivation in much of my research. Unfortunately, the experience of learning pronunciation at school of many adult learners whom I have taught has been negative, or their experience of using English has been tainted by their negative perception of their own accent. When memories influence an individual's perceptions of their own performance and thereby influence their motivation to learn, this is known as "attribution" (Weiner, 1976; 1980; 2010), and I have found through research and teaching experience, that there are many motivational levers which can be used to help to counter negative motivation.

In self-determination theory, Deci & Ryan (1980) explain how the two types of motivated behaviours, intrinsic and extrinsic, contribute to conscious, self-determined behaviours. They differentiate self-determined behaviours from automated behaviours in this early version of their the model. Intrinsic motivation provides individuals with the necessary "energy for decision making and managing motives", and is grounded in a "primary organismic need for competent, self-determined interactions with the environment" (Deci & Ryan, 1980: 35). Extrinsically motivated activities typically involve money-making for professionally active adults, exam results for school and university contexts, etc. These two concepts have been integrated into many motivation models. In this section, I will limit myself to a few concepts which I have found useful for explaining some of the data I have obtained in my own research, and for better understanding some of the difficulties various learners may have in improving their English pronunciation.

Bandura's social learning theory (Bandura, 1977a) is an extremely influential theory in the field of motivation and learning research, as it stresses the importance of efficacy beliefs and self-efficacy as a driving factor of motivation. Social learning theory suggests that there are three ways that people learn from each other; observation, imitation and modelling. Bandura's definition of imitation is the actual reproduction of observed motor activities, a definition which is particularly apt for an embodied approach to teaching pronunciation. Perhaps the most important concept in Bandura's social learning theory, and the one he is most remembered for, is self-efficacy. This concept is an explanation of how confident someone is feeling about the result of an action being successful or not, and how this may impact on their motivation to perform that action, as the following figure illustrates:



**Fig 3.2.3. Diagrammatic representation of the difference between efficacy expectations and outcome expectations (Bandura, 1977b: 193)**

As Bandura explains: “An efficacy expectation is the conviction that one can successfully execute the behavior required to produce the outcomes.” (Bandura, 1977b: 193). Negative efficacy expectations may be one of the reasons that many learners, and indeed teachers, fail to engage with learning pronunciation as effectively as they might. This is unfortunate, as there is a very large body of evidence showing that teaching pronunciation usually has positive outcomes, as we will see in chapter six.

Bandura explains that an individual can improve their feelings of self-efficacy, which has a positive effect on their motivation, through the “intervening influences of goal-setting and self-evaluative reactions” (Bandura, 1977b: 193). Firstly, goal-setting, most often associated with Locke (Locke 1968; Locke and Latham, 2006), can be a powerful lever in learning and teaching. In the research I carried out for my PhD (Frost, 2008), I administered a questionnaire to the 105 participants in the study, and the motivational construct which correlated most strongly ( $p$  values of less than 0.01) with other constructs I was interested in, was goal-setting, namely intrinsic motivation and self-efficacy. When teaching pronunciation to teenagers or to adults, this knowledge is extremely useful, and I have found that setting realisable short and medium-term goals can have a very positive effect on both motivation and learning outcomes. The second influence mentioned by Bandura in the quotation above is “self-evaluation reactions”. Self-evaluation may be formalised in teaching and in research with various self-assessment tools, from self-reporting, to questionnaires, to learners applying assessment criteria to their own performances. I have found self-assessment to be a particularly powerful tool for motivating students, although the experience can sometimes be a little painful. Participants in the ELLO study (Frost & O’Donnell, 2015) frequently used negative adjectives to describe their feelings on viewing videos of their oral productions at the beginning of the first year of their language degree courses, only to say how those negative feelings motivated them to improve. These students were setting themselves goals, and thereby helping themselves to regulate their own behaviour and take control of their own learning.

All theories and models must, to some extent, attempt to account for individual cognitive processes and interaction with the environment. Sociocultural theory, of which Piaget and Vygotsky were the

forerunners (Mahn & John-Steiner, 2012), is perhaps the best framework to understand the way interactions with the environment aids cognitive development. Understanding how concepts such as self-efficacy and goal-setting can positively impact motivation and learning can feed into better pedagogical practices.

### 3.3. Foreign-accented speech and listener judgements

We saw in the previous chapter to what extent L1 accent may act as a strong marker of identity within a language community, in that it enables other members of the community to make judgements about an individual. The same is true of course of L2 accented speech, or “foreign-accented” speech. Foreign accent has also been linked to other cognitive or affective judgements or inferences, based on perceptions associated with previously held individual and cultural beliefs. Accent has been referred to as “the face of language” by John Levis (Derwing & Munro, 2015: 139), and when listeners hear foreign-accented speech, they make conscious and unconscious judgements based on their accents. L2 users may have integrative reasons for “losing” their foreign accent (Bongaerts et al., 2000; McCrocklin & Link, 2016), or may be proud of their otherness, of which their accent is a marker (Cutler, 2014: 159). Hahn also asked the listeners to react to several statements regarding non-linguistic factors, linked mainly to the teaching assistant’s pedagogical abilities (Hahn, 2004: 213), and the results showed that the responses were generally more positive for the lecture with the correct stress patterns. Shiri Lev-Ari and Boaz Keysar investigated the credibility of foreign-accented speech in a study comparing the reactions of 30 L1 American English speakers to the veracity of random trivia statements uttered either in L1 American English speech, slightly accented speech, or heavily accented speech (Lev-Ari & Keysar, 2010). Participants rated statements uttered with a mild accent and with an L1 accent as being equally truthful, but the statements uttered with a heavy accent were significantly more likely to be rated as false. A perceived lack of credibility has been found by researchers in other contexts, such as Punjabi-accented American English (Dragojevic & Giles, 2016) and French, Swiss, Italian and English accents in German (Stocker, 2017). Alice Foucart and Susanne Brouwer (2021) also investigated the effect of foreign accents on moral judgements in in two different experiments, one comparing L1 Spanish speakers to British and Cameroonian accented speech, and the other comparing L1 Dutch to English-, French-, and Turkish-accented Dutch. Using classic moral dilemmas, this study showed that foreign accent influenced moral judgement, with an increase in utilitarian decisions for the Cameroonian- and French-accented speech when compared to the L1 Spanish or Dutch accents, respectively (Foucart & Brouwer, 2021: 1). This effect was significant, but not linked to comprehensibility, only to accentedness (Foucart & Brouwer, 2021: 7).



Elisabeth Zetterholm and Åsa Abelin (2019) investigated the reactions of listeners with various linguistic origins to the speech of Finnish-, French-, Arabic-, and Spanish-accented Swedish. The eleven evaluative dimensions included friendly, polite, helpful, happy, trustworthy, surprised, and energetic (vs. aggressive, uninterested, sad, and contemptuous). As with Foucart and Brouwer's study (2021), their results did not correlate with comprehensibility, nor even with a high degree of perceived accentedness. In this study, speakers were most positively evaluated if they had a high F0 variation, and most negatively evaluated if they had a more monotonous tone (Zetterholm & Abelin, 2019: 2167). A flat intonation pattern has also been related to negative listener judgements by other studies: Ron Thomson and Talia Isaacs (2021) investigated the effect of Mandarin- and Slavic-accented English on L1 Canadian English users' judgements in the areas of fluency, comprehensibility, friendliness, intelligence, and listeners' comfort. The results revealed that intonation played an important role in evaluations, which disadvantaged the Mandarin speakers, who had smaller F0 ranges. Pruned syllables and speakers' maximum F0 combined were strong predictors of both intelligence and interactional comfort ratings. F0 range also affected friendliness ratings in both groups (Thomson & Isaacs, 2021: 40). These findings corroborate previous studies, where flat intonation patterns have also been linked to other non-linguistic attitudinal judgements, such as boredom. In Okim Kang's (2010) study she reported that the undergraduate raters commented that Mandarin-accented international teaching assistants' "flat tone of voice made the speech not only frustrating but also extremely boring." (Kang, 2010: 310).

In the context of the French accent, Flege (1984) showed that it was possible for certain American English speakers to detect the French accent in a mere 30 milliseconds. Once a foreign accent has been identified, any negative judgments depends of course on the listener. A listener may have preconceptions about a professional competence or credibility based on strength of accent (Pettorino et al., 2012; Lev-Ari & Keysar, 2010), or may simply find the cognitive load of processing a foreign accent annoying (Tulaja, 2021). Irritation was also measured in Joan Fayer and Emily Krasinsky's (1987) study, and they linked it to "the communicative effect of errors" (Fayer and Krasinsky, 1987: 315). The original source of the irritation is accentness hindering intelligibility. Results showed that pronunciation and pauses contributed most to distraction in both groups of listeners, and that L1 Spanish listeners to Spanish-accented English were more annoyed than L1 English listeners (Fayer & Krasinski, 1987: 322-323). Cognitive load theory (Sweller, 1988), states that the mental effort required to solve problems increases as working memory is solicited by an increased number of problems to solve, and this leads to an increased cognitive load. Frequent pronunciation errors and unexpected hesitations are typically problems which contribute to linguistic processing (Martin, 1968; Phillips et al., 2022), and therefore increase cognitive load. Cognitive load is not only linked to working memory,

but also to long-term memory, and this is particularly evident with complex learning, defined by Van Meriënboer and Sweller (2005: 156) as learning to deal with “materials incorporating an enormous number of interacting elements”. Anything which increases cognitive load, is also very likely to impact motivation, and also memory (Feldon et al., 2019), and this can also affect learning.

In a study investigating the effects of prosody on learning novel words, Hadas Shintel, Nathan Anderson, and Kimberly Fenn showed participants images and played them novel words describing the pictures, either with congruent (correct, or normal) or incongruent (incorrect, or unexpected) prosody (Shintel et al., 2014). This was not an LX study, as L1 American English speakers and listeners were used, but the incongruent prosodic forms used, i.e. rising and falling intonation patterns, are typical of the sort of intonation error that French learners make (Herment et al., 2014). The team found that the prosodic form used did not have a bearing on the immediate recall tests, but when the participants were tested after a 24-hour delay, the words which had been spoken with a congruent intonation pattern were significantly more accurately recalled (Shintel et al., 2014: 1440). Results from this study suggest that listeners extract semantic information from prosody even when it is redundant, and that prosody can improve not only comprehension, but also recall (Shintel et al., 2014: 1437). The implications of this study for LX learning are important. If communication is to be effective, the information communicated by LX users will be better retained if the appropriate prosodic forms are used; in contexts such as Content and Language Integrated Learning (CLIL), or English Medium Instruction (EMI), this is particularly relevant.

In the field of English teaching, it has been shown that learners prefer to have “native speakers” rather than “non-native speakers” as teachers or teaching assistants (Hertel & Sunderman, 2009), and that they make their judgments as to the teacher’s L1 based on their accent. However, the students’ in Hertel and Sunderman’s study did not limit their judgments to the appropriacy of the “non-native” teacher to the teaching situation, they also made a number of other assumptions based on the teaching assistant’s accent. They assumed not only that they would learn better from the “native speaker” teacher, but also that the “native speaker” teacher also had better knowledge of the subject matter, and better teaching ability. I will address the native speaker debate more fully in chapter seven.

Finally, it is equally possible that speakers may themselves see the strength of their accent as a barrier to communication, or social integration, or that it may lead to other negative perceptions about them (Glusjek et al., 2011). Whether conscious or unconscious, listener bias based on a perceived foreign accent is a potential problem for all learners. As Oyama (1976) states: “One would hope that such considerations might diminish in importance as we become more enlightened in our linguistic

attitudes, but as long as they remain a social reality, we ignore them at our students' peril" (Oyama, 1976: 280).

### 3.4. Comparing and contrasting languages: transfer and associated issues

Language transfer, or crosslinguistic influence, may be defined as the influence of one language crossing into another language used by the same person. Lado (1957) is often credited with formalising the concept of linguistic transfer, which he explains as follows:

Individuals tend to transfer the forms and meanings, and the distribution of forms and meanings of their native language and culture to the foreign language and culture - both productively when attempting to speak the language and to act in the culture, and receptively when attempting to grasp and understand the language and the culture as practised by natives. (Lado, 1957: 3)

In cases of bilingualism, or multilingualism, any of a user's languages may influence the others. The most relevant form of transfer for LX phonology research and teaching is transfer from a language user's L1 or other previously learned languages into the target language. The resulting form of the L2, the "learner language", is often referred to as interlanguage (Selinker, 1972). This transfer may be negative, (errors may be introduced into the target LX from another language); this process of negative transfer used to be referred to as interference (Brière, 1966), and some authors still use this term (Crystal, 2008; Flege, 2021). Of course, not all language transfer is negative – If a learner's L1 has similar features to a target language, then similarity may ease the learning of that feature (Ringbom & Jarvis, 2009), and a teacher may use similar features as a starting point, etc. Transfer may occur in any of the area of reception or production, but we are particularly concerned with oral skills. As we will see in section 3.3 when we present some LX learning models, perception and production are cognitively associated processes: much as motor neurons can be triggered by observing a process (Ping et al., 2014), so speech production neurons are triggered when we listen to speech (Pulvermüller, 2005). Nonetheless, in the following sections, we will briefly explore some of the concepts connected with language transfer, starting with production and perception, then moving onto other transfer-related issues, and as always, relating the issues discussed to prosody.

#### 3.4.1. *Transfer in production*

The most easily noticeable form of transfer when a learner speaks, is a "foreign" accent. Crystal (2008) describes accent as follows: "The cumulative auditory effect of those features of pronunciation which identify where a person is from, regionally or socially" (Crystal, 2008: 3). Interestingly, Crystal's definition shares the onus of accent between the speaker and listener. The impact of prosody on

different constructs linked to foreign accent will be explored in chapter five, but since the 1990s, research-driven approaches to pronunciation teaching have increasingly stated the importance of prosody as a contributing factor to foreign accent in the field of English as an LX, and also emphasised it as an area of accent where change may be affected (Celce-Murcia et al, 1996).

Prosodic transfer at the level of production has usually been dealt with either in the context of prosody instruction in LX learning (e.g. Archibald, 1992; Carpenter, 2015; Chen & Mennen, 2008; Gabriel et al., 2012; Gabriel & Kireva, 2014; Ordin et al., 2011; Rao, 2011; White & Mattys, 2007), or in the context of socio-phonetics, by examining the prosody of immigrants in the language of their adopted country (e.g. Bordial, 2012; Deterding, 2001; Fagyal, 2010). Laurent Rasier and Philippe Hiligsmann investigated prosodic transfer of French learners of Dutch, and found evidence that prosodic transfer errors are cumulative (Rasier & Hiligsmann, 2009: 59) More recently, Anna De Meo and Massimo Pettorino's edited volume on interlanguage prosody provides a valuable contribution to the field, explores the question of interlanguage prosody in Italian from several different angles (De Meo & Pettorino, 2012). Chapter five will deal in more detail with the role of prosody transfer in the context of LX instruction studies.

Prosodic transfer in production can lead to misunderstandings with potentially serious consequences. For example, incorrect stress on numbers can cause misunderstandings between an air traffic controllers and pilots. Wilner and Feinstein-Whittaker (2007: 14), in a training manual for doctors, stresses the importance of clearly enunciating the difference between *15 mg* versus *50 mg*. Mompean, (2014) provides a detailed study of this "stress shift problem" in numbers in English and the misunderstandings this can cause. Julia Trippe and Melissa Baese-Berk (2019) draw attention to the problem of prosodic errors and their potential consequences in air traffic control, and highlight the importance of accurate prosody in aviation English (AE):

AE has a more restricted pitch range, is faster, and exhibits less variable vowel durations and more variable consonant durations than Standard English. These prosodic differences from Standard English may create difficulties for Aviation English users, and indicate inaccuracy in the assumption that attaining proficiency in conversational English is sufficient for proficiency in Aviation English. (Trippe & Baese-Berk, 2019: 30)

I will deal in more detail with studies which examine prosodic transfer in relation to intelligibility and language instruction in chapter six.

### *3.4.2. Transfer in perception*

In the early years of phonological theory, Polivanov (1931) wrote that we perceive sounds in another language through the sound system of our first language. Trubetzkoy (1939) used the metaphor of a

filter, suggesting that that inadequate production of L2 sounds had a perceptual basis, with the L1 system acting as a “phonological filter” through which LX sounds are perceived and classified. On the subject of “transfer of a native sound system”, Robert Lado also gave credence to the importance of perception, and mentioned “perception blind spots” when language users dealt with unfamiliar sounds (Lado, 1957: 11).

If, however, an LX contains features which are similar in nature to a listener’s L1, then it may be easier to perceive sounds and thereby access comprehension. Edensor (forthcoming) has found evidence that French speakers have fewer problems with regional varieties of English which are closer in rhythmical structure to French than others, for example Welsh-accented English, although participants were unfamiliar with this variety of English before the experiment. Iverson et al. (2006), after conducting experiments on the perception of rhythmic grouping of non-linguistic stimuli by Japanese and American listeners, found that their perception of rhythmical sounds was influenced by the rhythmical structure of their own respective languages. Clark and Garrett (2004) suggest that for LX learners whose L1 is also syllable-timed, it may be easier to perceive and comprehend a language or accent which is rhythmically closer to their L1. They performed a series of perception tests of foreign-accented speech on listeners who had received training in previously unfamiliar foreign accents, and demonstrated that their reaction times and perception noticeably improved, even after relatively short periods of instruction (Clarke and Garret, 2004).

Other authors concentrate on prosody in perception training to improve both perception and production in LX learning (Couper, 2022a), and in production work to improve perception (Cauldwell, 2013). Patel (2011), also places prosody at the centre of his OPERA model, a framework that explains how general non-language-specific musical activities can facilitate speech and language processing. The similarities he draws between the rhythm and melody of language and speech are at the heart of OPERA, and we will explore this framework more closely in chapter seven.

The extent to which the “phonological filter” outlined by the founding members of the Prague Linguistic Circle in the first half of the twentieth century operates, depends on the extent of the differences between the phonological systems of the two languages concerned. The concept of “stress deafness” (Dupoux et al., 1997; Dupoux & Peperkamp, 2001) was posited to account for difficulties L1 users of “non-stress” languages experience when learning “stress languages”. In their first study (Dupoux et al., 1997) the team conducted a series of four experiments involving French and Spanish participants and they found that French L1 listeners performed less well in perception tasks where they had to discriminate between trisyllabic items distinguished by “accent” (stress). They suggest that because French has no lexical stress, French listeners do not use stress as a factor when coding

information in their short-term memory after the acoustic information has been processed (Dupoux et al., 1997: 415). The particular difficulties encountered by French learners when learning a stress language has been outlined by Dupoux and his colleagues (Dupoux et al., 2008). Olle Kjellin uses the medical term for partial or slight deafness “amblyacousia”, (Kjellin, 1999: 380) when describing this phenomenon amongst adult learners of LX Swedish, which, he maintains, is “curable” by concentrating on prosody in an embodied approach when working on pronunciation with adult LX learners.

### 3.4.3. *Contrastive analysis and markedness*

Building on notions of transfer between languages, Fries (1945) stated:

The most efficient materials are those that are based upon a scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner. (Fries, 1945: 9).

It was Robert Lado (1957) who laid out a set of phonological principles for conducting this comparison, or what has come to be known as the contrastive analysis hypothesis. Lado echoes the ideas of Fries regarding relative difference of features in a learner’s first language and the target language as a predictor of ease or difficulty of learning:

the student who comes in contact with a foreign language will find some features of it quite easy and others extremely difficult. Those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult. (Lado, 1957: 2)

It may be assumed from the assertions of Fries and Lado that the degree of difference may have a predictive effect on how easy features may or may not be to learn in an LX. This is what Ronald Wardaugh (1970) called the strong version of contrastive analysis hypothesis, which he insisted would be “unrealistic and impracticable” (Wardaugh, 1970: 124). Along similar lines to Lado’s approach, Fred Eckman (1970) suggested that contrastive analysis could be revised to predict “directionality of difficulty”, in effect, a strong version of Lado’s hypothesis. Eckman suggests that markedness is the key to adding the missing component, that of “relative degree of difficulty”, which he explains as follows: “A phenomenon A in some language is more marked than B if the presence of A in a language implies the presence of B; but the presence of B does not imply the presence of A” (Eckman, 1977: 320). Eckman proposes the markedness differential hypothesis, and wishing to make his model as inclusive as possible, draws on the principles of universal grammar to refine markedness relations:

The areas of difficulty that a language learner will have can be predicted on the basis of a systematic comparison of the grammars of the native language, the target language and the markedness relations stated in universal grammar, such that,

(a) Those areas of the target language which differ from the native language and are more marked than the native language will be difficult.

(b) The relative degree of difficulty of the areas of the target language which are more marked than the native language will correspond to the relative degree of markedness; (c) Those areas of the target language which are different from the native language, but are not more marked than the native language will not be difficult. (Eckman, 1977: 321)

Markedness, a principle pioneered by Nikolai Trubetzkoy and Roman Jakobson during World War 2 (Eckman, 2008: 97), is based on binary oppositions. A marked feature is one which occurs more frequently in a language. Typological markedness is therefore a way of distinguishing linguistic representations between languages, as Eckman states:

an asymmetric, irreflexive and transitive relationship between linguistic representations across the world's languages, such that the presence of one structure in a language implies the presence of another structure, but not vice versa. (Eckman, 2008: 99)

If we take lexical stress as an example, it is strongly marked in English, and not marked at all in French, so (a) and (b) from Eckman's hypothesis as stated above clearly apply. The authors discussed thus far in this section mention only segmental features when applying the contrastive analysis hypothesis or markedness to phonological features, however Eckman (2004) discusses prosody at length, stating:

Stress has been the most studied aspect of L2 prosodic structure+ In addition to the above investigations showing the role of UG parameters in the SLA of stress, there has also been work attesting to the effect of L1 transfer. Thus, the findings have turned out to be similar to the results seen in other L2 phonological domains—namely, the L1 plays a significant role in determining IL [interlanguage] stress, and at times the learners have constructed an IL stress system that derives from neither the L1 nor the L2. (Eckman, 2004, 537)

Eckman (2004) cites Pater's (1997) study detailing the difficulties French learners experienced with learning stress patterns in English, "missetting" the parameters of word headedness and directionality, and obtaining a result which was neither French nor English.

#### *3.4.4. Optimality Theory*

Another framework by which languages may be compared is optimality theory (Prince & Smolensky, 1993). It is a way of looking at the organisation of a language by examining its grammar via a system of ranked constraints (which may be seen as rules) and it is often used to better understand language transfer (Eckman, 2004). It is not technically a theory, in that it does not make falsifiable predictions, but it is a model which may help us to understand the distribution, if not the nature, of certain metrical features of language. One of the sets of constraints which optimality theory includes is markedness. There are many "constraints" in each set, and example for a markedness constraint being the Weight-to-Stress Principle, which states "Heavy syllables are prominent in foot structure and on the grid. By this principle, the trochaic group (L3 H) is subpar because it puts a heavy syllable in a weak position" (Prince & Smolensky, 1993: 56). Optimality theory has been much criticized, not least because of its

attempt at universality, a goal which, as we saw with Pike's stress-timed vs. syllable-timed dichotomy, is an ambitious aim, given the number and diversity of languages in the world. In a footnote to the Weight-to-stress principle, which was based on an analysis of only a handful of languages (including Latin), the authors do concede that they have to adapt certain other constraints so as not to violate them, adding "we incorrectly predict antepenultimate rather than penultimate stress in light-syllabled words. The foot structure of languages like Hindi is, to say the least, incompletely understood." (Prince & Smolensky, 1993: 56, footnote 33). Although optimality theory has its weaknesses, it is still responsible for a great deal of ongoing research, for example Maria Gouskova (2015) examined the theory applied to phonology, paying special attention to questions of typology, learnability, acquisition, and variation. She concluded: "it is still a vibrant theory with many directions for development" (Gouskova, 2015: 564).

#### 3.4.5. *Functional Load*

Functional load theory was another product of the Prague School of linguists which can be used to compare features between languages. Crystal defines functional load as follows:

The use made of a linguistic contrast in a system is sometimes referred to as its **functional load** or **yield**. The term is usually used with reference to phonology, where in English, for example, the contrast between /p/ and /b/ would be said to have a higher functional load than between /a/ and /e/: the former contrast distinguishes many minimal pairs, whereas the latter contrast distinguishes only a few. (Crystal, 2008: 201)

As Crystal's example demonstrates, most of the work on functional load theory concerns functional load of segments, especially vowels. However, in English, lexical stress and prominence at the tone unit level have a very high potential to distinguish meaning, and therefore according to this theory, a very high load. Functional load theory has also been used to explain the relative importance of prosodic features, especially prominence (Vogel et al., 2015). Some authors have used it as a factor to help decide where to spend time when teaching pronunciation (Brown, 1988).

Munro & Derwing (2006) also suggest using this theory to choose which items to work on when teaching pronunciation, and although they focus on consonants in their study, they acknowledge that attention to prosody is well justified, and that a functional load analysis must encompass "all aspects of speech including consonants, vowels, prosodic factors, voice quality, and other more general speech characteristics, while taking into account the context of communication" (Munro & Derwing, 2006: 530). Sewell (2021) calls for more research on the functional load principle in pronunciation teaching choices, however warns that a high functional load should be weighed against learnability (Sewell, 2021: 4). Alnafisah and his colleagues, in a study which focused on segmental errors and functional



load theory, also highlight the usefulness of this theory, especially when comprehensibility is taken into consideration, concluding:

It is rare in language teaching research to find theories that have relatively direct implications for classroom practice. The evidence from this study and others that have preceded it indicates that functional load is one of the rare instances that can give teachers clear guidance about how to help learners improve the comprehensibility of their spoken language, not by adding additional features to work on, but by targeting only those features that are likely to make a difference to listeners. (Alnafisah et al. 2022: 12)

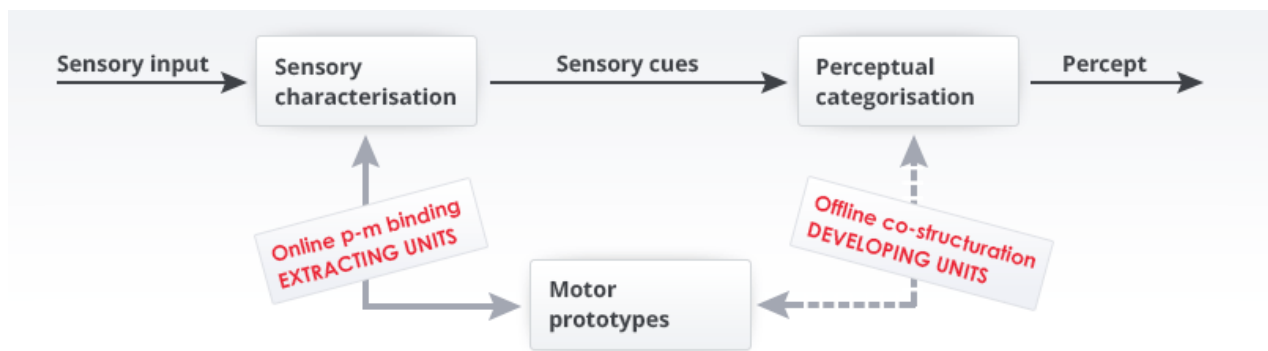
### 3.5. From speech perception and production to memorisation & language acquisition

In this section, I would like to present some theories and models which are particularly relevant to an embodied cognitive approach to LX pronunciation teaching and research, Firstly, I would like to briefly describe the motor theory of speech perception which, along with the research on visual activity and stimulation of the motor centres of the brain, has important implications for teaching pronunciation. Secondly, I will present two versions of Levelt's model of language production, and why it is particularly relevant to an embodied cognitive approach to pronunciation instruction. Section 3.5.3 will deal with some influential LX speech models, and finally, in section 3.5.4. I would like to present a Brian MacWhinney's (2005) unified competition model of language acquisition.

#### 3.5.1. *Speech perception: motor theories*

Several models from the 1950s to the 1980s can be described as motor theories, in that they consider language perception to be an active process, implying that perception involves the stimulation of motor activity in the brain. This may be what Daniel Jones (1918) referred to as the "inner voice". Alvin Liberman & Ignatius Mattingly (1985) suggest in their revised motor theory that "the objects of speech perception are the intended phonetic gestures of the speaker" (Liberman & Mattingly, 1985: 2), by which they mean that the brain relies on the same areas for these gestures when listening as when speaking.

The PACT (Perception-for-action-control) model is a perceptuo-motor model proposed by Jean-Luc Schwartz and his colleagues at the University of Grenoble Alpes (Schwartz et al., 2012). This multisensory model provides a framework for connecting perceptual shaping and motor procedural knowledge in speech. The model is illustrated in figure 3.5.1:



**Figure 3.5.1. The PACT model of speech perception (Schwartz et al., 2012)**

This model proposes a theoretical framework which connects perceptual shaping and motor procedural knowledge in multisensory speech processing in the human brain (Schwartz et al., 2012: 337). In the PACT model, the route to procedural knowledge involves a “perceptuo-motor link” which contributes to structuring perceptual categories in reference to their motor content (Schwartz et al., 2012: 348), which is the principle of motor theories of perception in general.

As Carol Fowler points out, the activation of motor areas of the brain during perception is not limited to language:

Motor recruitment occurs generally in perception and cognition, including in language perception and comprehension. This is likely because life in the eco-niche is pervasively perceptuo-motor in nature, and animals, including humans, are adapted to that kind of life. (Fowler, 2016: 178)

The motor theory of perception, which Fowler (2016: 181) now calls “plausible, even mundane”, or at least the strong links between perception and production, have been confirmed by a vast body of work in neuroimager, especially PET and fMRI (Price, 2012). Sakai and Moorman’s meta-analysis of perception training research provides a comprehensive review of the last 25 years of L2 perception training studies that test for effects in production, and the results clearly indicate that “the two modalities are connected, insomuch as training the perception of L2 sounds can induce positive change in the productive mode as well” (Sakai & Moorman, 2018: 199). This research has clear implications for pronunciation teaching.

### *3.5.2. Speech production: the role of feedback*

Willem Levelt’s (1989) model has been highly influential in language production research. Its three stages, involving the conceptualisation, formulation, and articulation of language, illustrate the

movement from intention, through various cognitive processes, to the physical articulation of speech, as we can see in figure 3.5.2a:

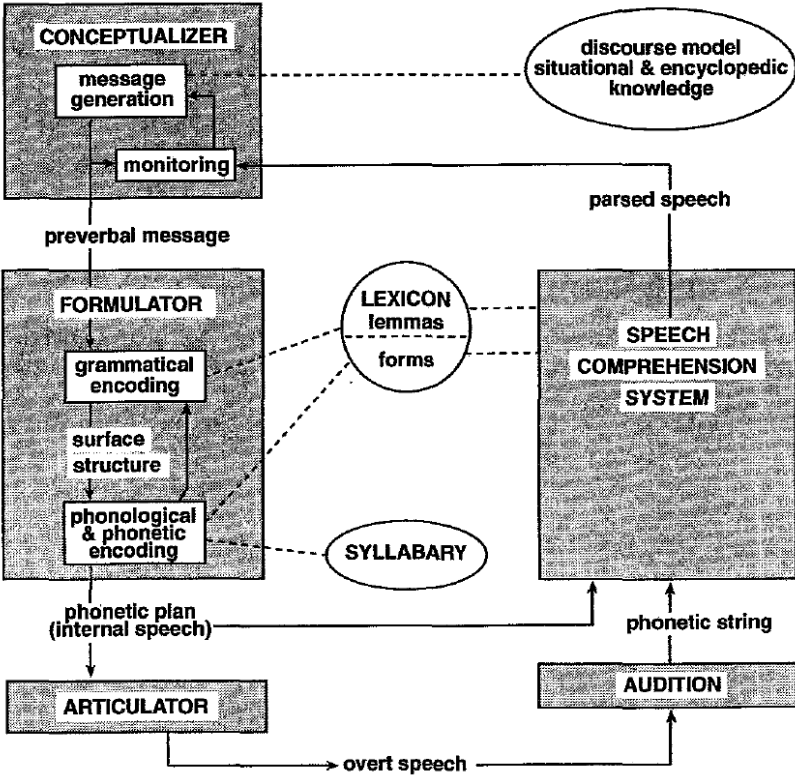
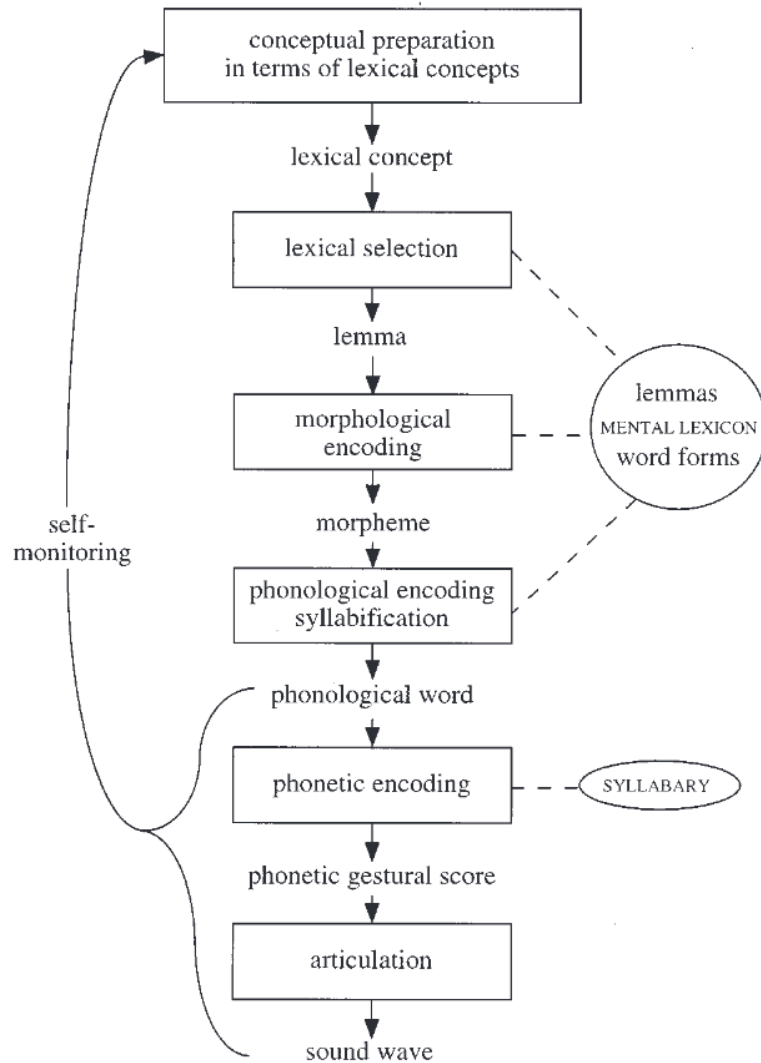


Figure 3.5.2a. Levelt’s model of speech production (Levelt, 1989: 9)

Levelt’s original model suggests that monitoring plays a role at the conceptualisation stage, and is influenced by the listener, which has clear implications for feedback in the context of teaching and learning languages. Levelt provided more detail on the processes involved in lexical selection in a later paper (Levelt et al., 1999), as figure 3.5.2b illustrates:



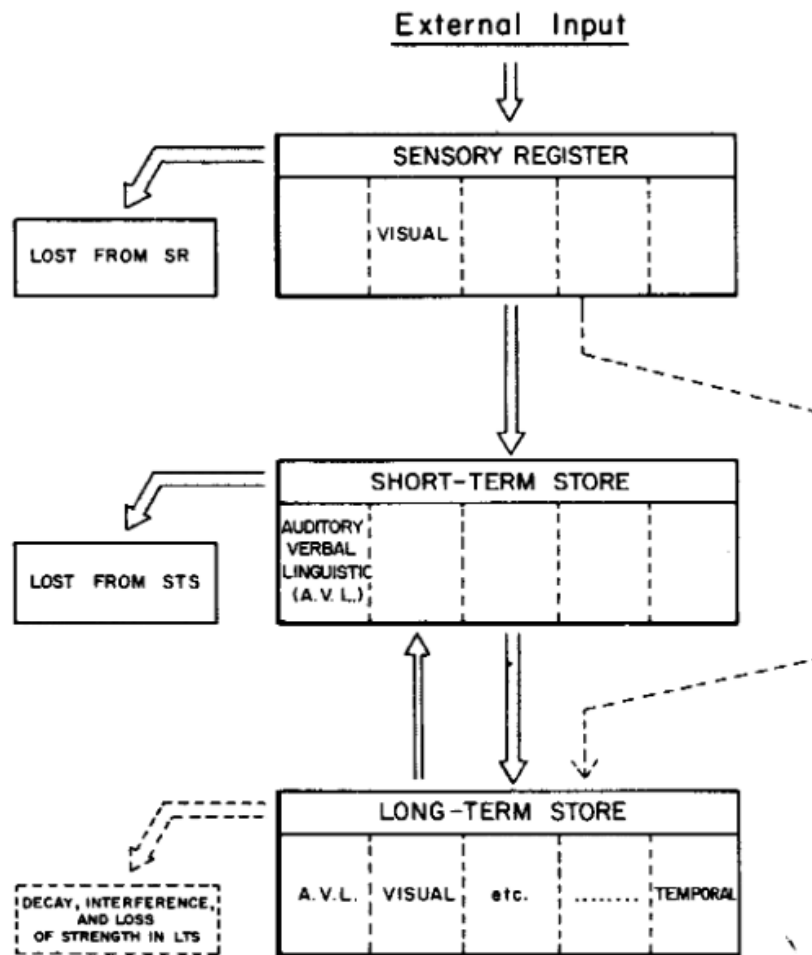
**Figure 3.5.2b. Levelt's model of speech production (Levelt, 1999: 3)**

The most interesting refinement which this iteration of Levelt's model provides is the level at which monitoring occurs. No longer confined to the level of conceptualisation, in terms of pronunciation, monitoring occurs after phonological encoding, at the *phonetic* level. In the context of LX-accented speech, the physical process of articulation is what allows both speaker and listener to be aware of the features which contribute to accent. By intervening at this physical level, where speech is embodied, teachers can help learners to develop strategies which can improve the phonological and phonetic encoding of selected words.

### 3.5.3. *Memory, multimodality, and the role of haptic information*

It is not my intention here to present a complete review of our understanding of the human memory and learning. I would like, however, to highlight the multimodal nature of some of the most important

models of memory, and in particular, the role of haptic information in memory. The most influential model of human memory in the twentieth century was proposed by Richard Atkinson and Richard Shiffrin (1968). Following George Sperling's (1963) study, which led him to suggest an immediate memory for storing visual information as a component of short-term memory (Sperling, 1963: 21), Atkinson and Shiffrin suggested a three-stage model, represented in figure 3.5.3a:



**Figure 3.5.3a. The structure of the memory system (Atkinson & Shiffrin, 1968: 93)**

Atkinson and Shiffrin's model involves three stages, and begins with a multi-sensory register – necessarily, since all senses are involved in the perception of information which is encoded into the short-term and long-term memory store. The short-term store, which the authors suggest, based on available evidence, represents around 15-30 seconds, contains auditory, verbal & linguistic information, which Atkinson and Shiffrin treat as a whole in their model, explaining that it is not easy to separate these three functions (Atkinson & Shiffrin, 1968: 92).

For information to be memorised, it must first be perceived and encoded. Alan Paivio's dual coding theory (Paivio & Csapo, 1969) deals with the issue of perception via different channels, or modes of

input which Atkinson and Shiffrin’s model does not deal with. In this model, verbal stimuli and nonverbal stimuli are considered as being processed separately, as the diagram in figure 3.5.3b represents:

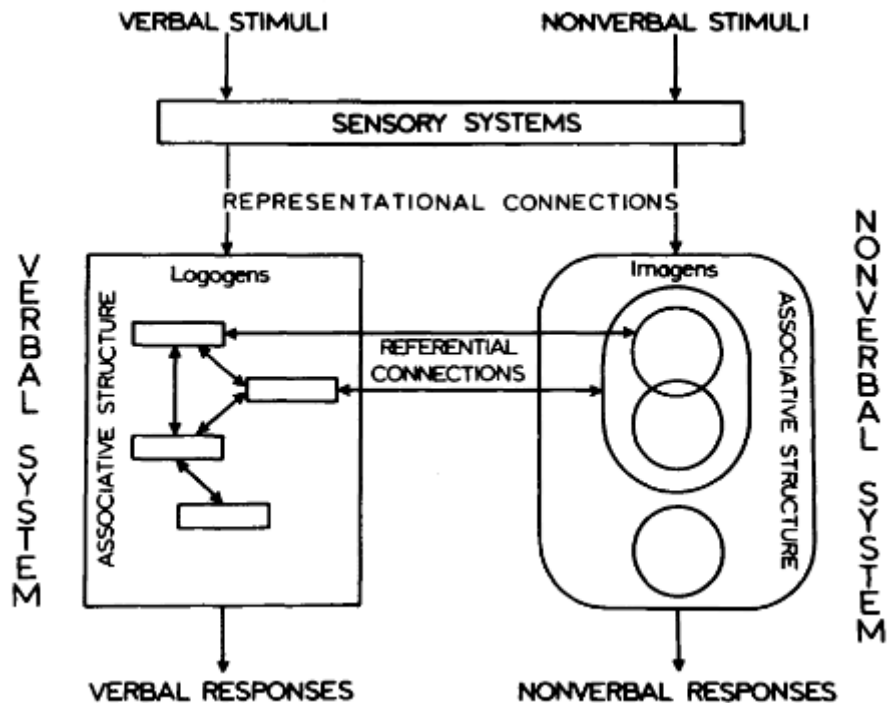
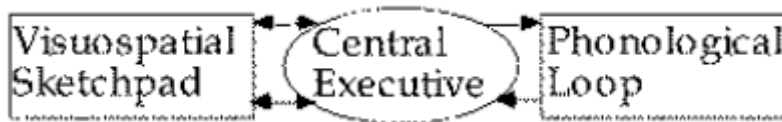


Figure 3.5.3b. Verbal and nonverbal symbolic systems of Dual Coding Theory (Paivio, 1986: 38)

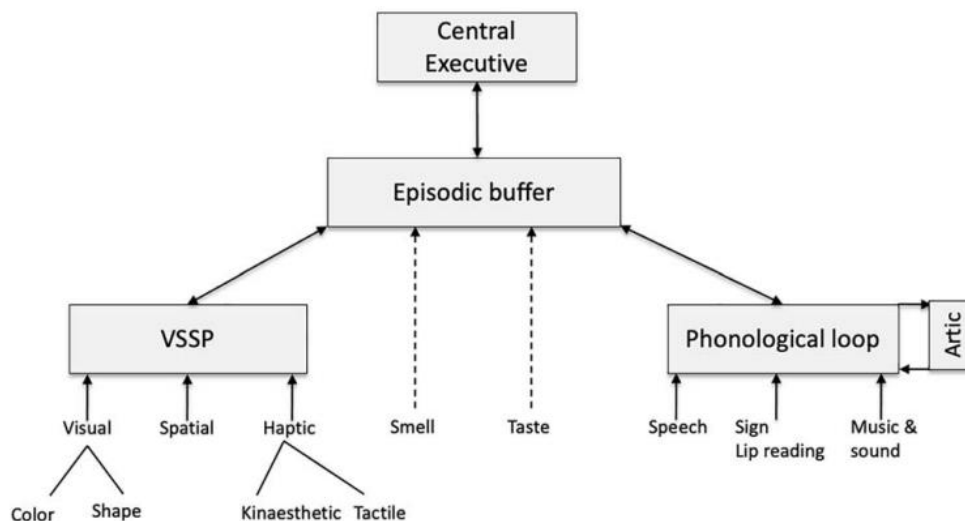
Paivio’s model is not a model of memory, but as a multiple coding theory, it emphasises the fundamental importance of the verbal/nonverbal contrast in coding information prior to learning – information which is then accessed from memory. As Paivio (1991) states, the sensorimotor side of the theory had often been overlooked, and this model does allow for different modalities of classes of events which make up the nonverbal system, and these include haptic, or “tactual and motor feedback” (Paivio, 1991: 257). Dual coding theory also influenced a range of multimodal approaches in education, involving verbal and nonverbal stimuli.

The second stage of Atkinson and Shiffrin’s model – the short-term store - was addressed by Baddeley and Hitch (1974), who proposed a model of “working memory” as an alternative to the “short-term memory” of the Atkinson-Shiffrin memory model. The notion of multimodality is also central to the Baddeley and Hitch model, since a visual “sketchpad” and a “phonological loop” support the attentional control system, called the “central executive”, as pictured in figure 3.5.3c:



**Figure 3.5.3c. The three-component model of working memory proposed by Baddeley and Hitch (1974) (Baddeley, 2003: 191)**

The concept of the phonological loop, and its role in the “subvocal rehearsal system (Baddeley, 2003: 191), as a means of reinforcing memorisation has clear implications for pronunciation instruction. However, as Susan Gathercole (2008: 43) suggests, this stage may constitute a “bottleneck” for learners. It also represents a stage in the process of memorisation where a multimodal approach, and the use of embodied activities can help to encode neurophysiological data in a more durable fashion. The encoding of multimodal data into a coherent unitary whole was not accounted for in Baddeley’s model, and the addition of the “episodic buffer” (Baddeley, 2000) was an attempt to rectify this. The episodic buffer enables experiences perceived and encoded through different channels to be chunked together and experienced as single objects or events. In light of the advances in neuropsychology, Alan Baddeley and his colleagues have reviewed their previous model of the working memory (Baddeley & Hitch, 1974; Baddeley, 2003). Their current model (figure 7.3b) emphasizes the importance of different types of physical inputs, and their combined effect along with the phonological loop and the visuo-spatial sketchpad in memorising information:



**Figure 3.5.3d. Baddeley’s current model of memory illustrating the flow of information from perception to the episodic buffer (Baddeley, 2021: 870)**

The models of memory in this brief review are all multimodal in nature, and they evince an increasing importance given to haptic information and its role in the process of memorisation. The use of physical

movement in learning has been shown to have a positive effect on memory: since 2010, over 150 journal articles and books have been published on this phenomenon, and its effect is, by the standards of psychology, considered very robust (Roberts et al., 2022).

Multimodal approaches to learning and teaching are certainly not recent, for example, Bloom's taxonomy (Bloom et al., 1956) includes three domains: cognitive, affective and psychomotor. However, as the amount of research and practice on the psychomotor domain was so limited at the time, Bloom and his colleagues did not write a third handbook corresponding to this domain (Bloom et al., 1956: 7-8), although several competing taxonomies for this domain have been proposed since (Gogus, 2012).

#### *3.5.4. Models in LX acquisition research*

LX acquisition models may be placed into three categories according to how they explain language *perception*. Firstly, those where the L1 influences the perception of the acoustic signal, i.e. different acoustic cues are weighted differently. Alexander Francis and Howard Nusbaum (2002) refer to these as “dimensional warping” models (Francis & Nusbaum, 2002: 350). Iverson & Kuhl (1995) use the metaphor of a “perceptual magnet” in their native language magnet model: “The [native language magnet] model predicts that the re-structuring of perceptual space that assists infants in the acquisition of phonology can adversely affect adults' subsequent ability to perceive foreign-language distinctions” (Iverson & Kuhl, 1995: 560). A second category of model suggests that perceptual difficulties are due to the similarity to abstract phonetic features (representations of phonetic features). The best-known model of this type is probably the Speech Learning Model, or SLM (Flege, 1995). Lado (1957) made similar suggestions, but using contrasting phonological, rather than phonetic features to explain perceptual difficulties. The third type of model holds that perceptual differences are due to interference from L1 categories or prototypes; the perceptual assimilation model or PAM (Best, 1996) is an example of this, providing four possibilities of assimilating features in the LX or not based on the similarity to phonemes in a speaker's language or languages learned from infancy. There are important teaching implications for these models, for example the PAM and SLM models imply that raising awareness to L1 and L2 differences would improve learners' chances of establishing new phonological categories in their own repertoires (Best and Tyler, 2007: 30).

As we saw in chapter one, prosodic features are among the first aspects of language we acquire. Few LX acquisition models thoroughly address the issues of age and transfer, however Flege's revised SLM (SLM-r) model (Flege, 2021) re-evaluates the question of age. While it is certainly true that most LX learners speak with a foreign accent, the degree of this accent cannot simply be explained by age, as



numerous critics of Flege's original model have pointed out (Walley, 2007). In his SLM-r model, Flege has carried forward the idea of phonetic categories, but this time, he states "L2 learners of any age make use of the same mechanisms and processes to learn L2 speech that children exploit when learning their L1" (Flege & Bohn, 2007: 14). He also concedes that there may be changes in the way that we all learn and lose language in both our L1 and other languages over the course of our lives, both as a normal part of brain development, and in the case of neurodegenerative conditions (Flege & Bohn, 2007, 14-15).

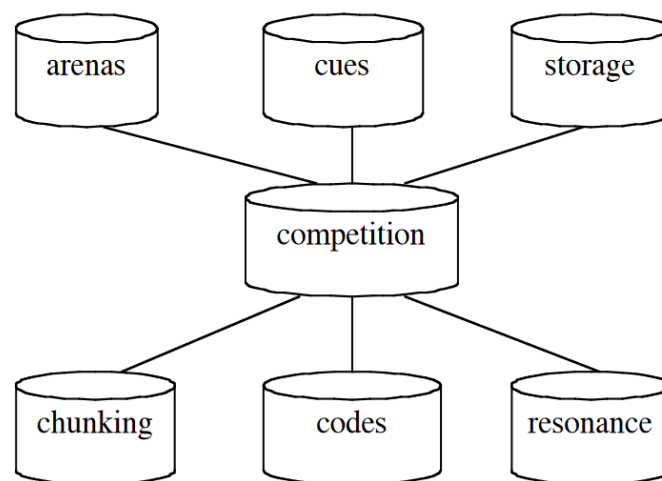
### *3.5.5. Towards a unified model*

The models of the 1990s and earlier presented in the previous sections either do not mention prosody at all, or mention it very little. Furthermore, as we have seen, the issue of language transfer is extremely relevant to LX phonological acquisition, particularly in adults. Krashen (1985) suggested that adults have two separate and independent systems for developing ability in second languages, subconscious "language acquisition" and conscious "language learning" and integrated this concept in his monitor model. This model has been highly influential in language teaching, and is one of the models which I discuss with trainee teachers, and the acquisition/learning distinction is one I find particularly useful. Recently, with the development of neural networks and artificial intelligence, there are working prototypes of models for artificial intelligence applications which account for, and allow natural language understanding and generation (Dong, 2019), however such models are of limited use to linguists and language teachers researching language acquisition and LX learning inside and outside the classroom. A more complete model for LX researchers and teachers would take account of the features mentioned so far in this section: the interplay between languages, physical and cognitive issues, individual differences and environmental factors - and not simply a narrow definition of "linguistic" factors. The model I will present is both simple and extremely comprehensive, and considers language users from an embodied cognitive approach.

The beginnings of this model were fairly modest: Brian MacWhinney and Elizabeth Bates (MacWhinney & Bates, 1982; 1989; MacWhinney, 1987) suggested a competition model to explain of acquisition of form-function relations in L1 development, based on the morpho-syntax of a handful of European languages. In its early form, this model had six tenets, the first of which concerned "channel limitations" (the constraints on the acoustic articulatory mechanisms involved in perceiving and producing four kinds of "signals"): "lexical items, word-order patterns, morphological markings and intonational contours", the last of which presumably includes prominence at both the lexical and sentence levels, (Bates & MacWhinney, 1982: 190).

Since language transfer is an integral part of LX learning, it played an increasingly important role in the model as it developed (MacWhinney, 1992). Transfer is not seen as entirely negative, but of course it often results in errors, and the treatment of these errors by a learner or in class are important for MacWhinney, who states that the competition model is most in accordance with a list of pedagogical principles. This list includes learning in context, repetition in early instruction, audio-active comparative work and CAPT, integrating phonological and grammatical instruction, considering positive and negative transfer, dealing with difficult forms and errors in reception and then moving onto production activities, and encouraging learners to monitor and self-correct (MacWhinney, 1992: 384-385).

MacWhinney also recommends the incorporation of metacognition regarding phonological transfer and the gradual movement from controlled to freer practice throughout the learning process, and these are principles which we will present in more detail in chapter seven. Over the three decades since the first version of the competition model, and drawing on studies in 18 languages with children and adult learners in monolingual and LX learning contexts<sup>9</sup>, the model has been refined to its current form, the unified competition model, as illustrated in fig 3.3.1:



**Figure 3.3.1. The Unified Competition Model (MacWhinney et al., 2005: 50)**

At the core of the model is the processing system which selects between various cues. The arenas are the levels of phonology, morphosyntax, lexicon and conceptualisation. Cues are linguistic signs, and storage refers to the long and short-term memory processes. Chunking is segmentation of speech into prosodic units, and is a process which is part of perception and production. The term “codes” involves a language user switching between languages, dialects, accents or register, depending on the needs of

---

<sup>9</sup> 142 studies are listed in MacWhinney, 2005.

a given situation, and finally resonance, the latest addition to the model, is defined by MacWhinney as follows:

resonance is based on the repeated coactivation of reciprocal connections. As the set of resonant connections grows, the possibilities for cross-associations and mutual activations grow and the language starts to form a coherent co-activating neural circuit. (MacWhinney, 2005: 61)

MacWhinney states that resonance “seeks to relate the Competition Model to research in the area of embodied or embedded cognition, as well as newer models of processing in neural networks” (MacWhinney, 2005: 51). Regarding transfer in articulation, MacWhinney states that for an L1 learner, the major challenge is production, whereas in older L2 learners, it is overcoming the effects of negative transfer. Positive transfer of certain articulatory features may initially aid communication, but may lead to fossilization. Older children may rely on neural flexibility to reinforce articulatory skills, as they would for other motor abilities, such as sports, but for adults, with careful training, MacWhinney notes that progress can be made, and that resonance calls for metacognition “to reinvigorate a motor learning process that runs much more naturally in children and adolescents.” (MacWhinney, 2005: 56), stating:

To permit the growth of resonance in L2, learners must apply additional learning strategies that would not have been needed for children. These strategies focus primarily on optimization of input, promotion of L2 resonance, and avoidance of processes that destroy input chunks. (MacWhinney, 2005: 55)

MacWhinney also highlights the advantages and the dangers of technology, which can increase learners’ access to input, but can also lead to isolation (MacWhinney, 2018). He draws attention to the importance of a contextualised curriculum, and active participation in the environment (MacWhinney, 2018: 304). MacWhinney’s Unified Competition Model is grounded in a framework of embodied cognition, and has some important pedagogical implications for the acquisition of phonological features.

As it is based on competing cues at the level of perception, and phonological, lexical, morphological and syntactic processes at the level of production, this competition results in the transfer of L1 features into the target language. For a teacher or a researcher to know which transferred features to focus on, a study of the languages concerned is essential, and in the next chapter, we will examine the features of French and English, paying particular attention to prosody and to prosody-related issues.

## CHAPTER FOUR

### A contrastive phonological analysis of French and English

In this chapter, we will present, compare and contrast the phonological features of English and French, paying particular attention to the prosody of these two languages, and to features which are influenced by prosody. I presented Lado's contrastive analysis hypothesis (1957) in the previous chapter, and how this may help language teachers to prioritise phonological features for learners. Contrastive analysis has its critics (Crystal, 2008: 112), and I do not suggest that a mere contrastive analysis of English and French would suffice for establishing learning objectives, however it is a necessary starting point. When combined with other approaches, such as the markedness differential hypothesis (Eckman, 1977), functional load theory, optimality theory (Prince and Smolensky 2004), the intelligibility principle (Levis, 2005), and learnability (Heine et al., 2015), a contrastive phonological analysis can help teachers to define learning objectives with their learners, and can help researchers refine objectives and measure learner progress. The contents of this final chapter of the first part have shaped my own teaching and research since my first contact with French learners of English.

In the handbook for language teachers *Teaching English Pronunciation*, Joanne Kenworthy (1987) provides a series of pronunciation features for a selection of groups of learners according to their L1, and established priorities based on intelligibility. In the introduction to this part of her book, she states "[i]n general, the areas of rhythm, word stress, and sentence stress are high priority areas for all learners" (Kenworthy, 1987: 123). Based on the research that I present here, Kenworthy's assertion is valid. In this chapter, prosodic features guide the contrastive phonological analysis of the two languages which interest us: firstly, we will examine stress, rhythm and intonation, along with the question of tonicity, or focus, then the vowels of French and English, and finally consonants.

Finally, it is important to distinguish between production and perception in such an analysis. Meaning is a co-construction between speaker and listener, but so are the features which make up prosody itself. Prosodic features are a result of measurable acoustic realities (variations in F0, intensity, duration and formant structure), and how they are perceived by the listener. In a cross-linguistic context, the effects of language transfer that I discussed in sections 3.4.1 and 3.4.2 are therefore as important as acoustic measurements of prosodic variables. This is an important factor in the concept of isochrony, which as we saw in section 1.1.4, is as much a perceptual phenomenon as it is an acoustic reality. It will become clear in section 4.1.5 that listener perceptions of differences in intonation between English and French are also as important as measurable acoustic features.

## 4.1. Stress and rhythm

Following Pike's (1945) description of stress-timed and syllable-timed languages, French was described as being a "syllable-timed language" (Abercrombie, 1967). I do not wish to revisit at length the dichotomous concept of isochrony as a classification principle for all the world's languages into stress-timed or syllable-timed languages (Pike, 1945), however, there are undoubtedly some important differences between the metrical structures of French and English, and in the two languages' rhythms. Most authors agree that the isochrony dichotomy is too simplistic (Roach, 1982), but that stress has a more or less structuring role in languages (Cruttenden, 1997). Arvinati (2009) calls for a relative, rather than an absolutist approach to rhythmical analysis of languages, and that is what we will set out to do in this chapter, beginning with word stress.

### 4.1.1. *Stress at the word level in English*

Firstly, as we saw in chapter one, English has a "stress accent", and certain syllables are given prominence (Cruttenden, 1997: 10). This means that any or all of the four acoustic cues to prominence may be used to differentiate a stressed syllable from the surrounding syllables. As we saw in section 1.1.2, stress in English exists at the word level or at the sentence level, and perhaps the most important difference between English and French is that English has word stress and French does not. Word stress may be described as syllable prominence at the word level (Arnold, 1957: 225; Cruttenden, 1997: 7). Okim Kang and her colleagues describe word stress in terms of the four perceptual cues of the acoustic cues to stress which we saw in chapter one:

Word stress, synonymous with lexical stress, refers to the syllable of a word which is louder, longer, and higher in pitch than the other syllables of that word (e.g., *uniVERsity*, *STUdent*, *proFESSor*, etc.). (Kang et al., 2022:45)

English has movable stress (Crystal, 2008: 455), which is to say that the stress does not always fall in a fixed and predictable pattern, for example on the last syllable, or the penultimate syllable, as it does for example in Spanish or Italian. Instead, the patterns governing stress placement in English are complex, and were most comprehensively laid out in Chomsky and Halle's *Sound Patterns of English* (1968). Stress has the highest functional load when it can distinguish the meaning between two lexical items, as does contrastive stress, which is described by Levis and Muller (2018) as a high value feature. Prominence typically indicates new and important information, and this is also true of lexical stress in English. At the word level, stress in English provides useful semantic, syntactic and morphological information. For example, English tends to stress the root form in words preceded by prefixes, such as

*a*RR**EST**, *b*e**Q**U**EST**, etc. At the syntactic level, word stress can differentiate between word pairs such as *TR*AN**S**fer (noun) and *trans*FER (verb). In compound nouns, the stress on the initial element provides semantic information and shows lexical cohesion (*BL*ACKbird, as opposed to *black* BIRD). The stress shift from the word *Universe* to the word *uni*VER**s**ity with the addition of the strong ending provides additional morphological information. On the perceptual level, word stress is therefore a very important feature in English for accessing meaning. In addition, errors in word stress placement or in the correct use of all of the acoustic cues to stress at the production level can contribute to intelligibility problems, as we will see in chapter six.

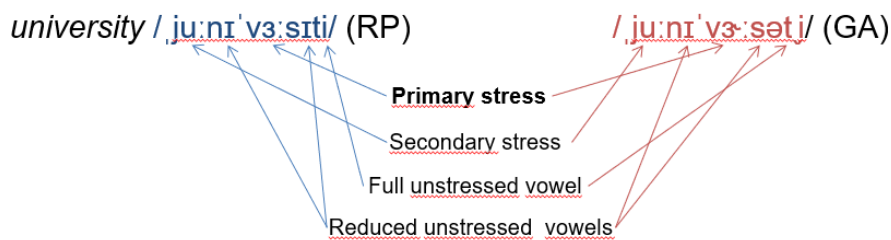
#### 4.1.2. Degrees of stress, and reduction in English

Although prosody is part of suprasegmental phonology, as opposed to segmental phonology, there is necessarily a certain amount of crossover into what are generally thought of as segmental features. Suprasegmental features are by definition larger than single phonemes. They are “properties of stretches of speech” (Pennington & Rogerson-Revell, 2019: 3), and so they necessarily affect the segments of speech. Regarding English prosody, as we saw in section 1.1.5, phonemes (particularly vowels), are often greatly affected by stress: when a syllable in English is stressed or unstressed, the four acoustic cues and their perceptual cues affect the vowel which composes the nucleus of that syllable. The nucleus of a stressed syllable will therefore have a **full vowel**, and an unstressed vowel may **reduce**, almost to the point of imperceptibility, as may completely disappear – a phenomenon known as schwa deletion (Hooper, 1976). In English, stressed vowels will fall into one of two categories, either tensing vowels /eɪ ɑː iː aɪ əʊ (j)uː/, or lax vowels /æ e ɪ ɒ ʌ ʊ/ (Ginésy, 1995). The stressed vowels are sometimes referred to as *fortis* and the unstressed vowels as *lenis*, following Arnold’s classification (1957).

A syllable which receives secondary stress (noted by a small low vertical diacritic), or is “lightly stressed” (Celce-Murcia et al., 1996: 132), or “partially stressed” (Cooper-Kuhlen, 1986: 31) does not reduce, but is nonetheless distinguishable from the primary stressed syllable. **Reduction** of an unstressed vowel in English usually results in schwa (/ə/), or sometimes /ɪ ʊ ɪə ʊə/. Reduced syllables may therefore be distinguished from partially reduced syllables, for instance: of the pair *postman* (UK English) – *mailman* (US English), only *postman* has a reduced vowel /pəʊstmən/ [vs. /meɪlmæn/]. This difference can be expressed: 'postman, 'mail,man (Cooper-Kuhlen, 1986: 31). For most authors, four degrees of stress are sufficient for analyses. Cruttenden (1997) distinguishes four degrees of stress for English as follows:

1. PRIMARY STRESS (or PRIMARY ACCENT), involving the principal pitch prominence in the intonation-group.
2. SECONDARY STRESS (or SECONDARY ACCENT), involving a subsidiary pitch prominence in the intonation-group.
3. TERTIARY STRESS, involving a prominence produced principally by length and/or loudness (and hence we cannot refer to this as 'tertiary accent' because the term 'accent' has been reserved for pitch prominences).
4. UNSTRESSED (the term UNACCENTED covers both (iii) and (iv)) (Cruttenden, 1997: 18)

An example of the four degrees, or levels of stress in English is the word *university*, as we can see in figure 4.1.2:



**Figure 4.1.2. The four degrees of stress in British (RP) and American (GA)**

In figure 4.1.2, the word *university* has the same stress pattern in British English and American English: the third syllable receives primary stress, the first receives secondary stress, and vowels in the second and fourth syllables both reduce to /ɪ/ in British English (RP), and to /ɪ/ and to schwa respectively in American English (GA). The final vowel does not actually receive tertiary stress, but is partially reduced, and maintains some tension due to its final position (Wells, 2012). The stressed syllable may also be marked in English by increased energy on the initial consonant (which may manifest as aspiration) or on the initial vowel (which may manifest as a glottal stop) although these “secondary cues” are not obligatory (Couper-Kuhlen, 1986: 20). It should also be noted that reduction in certain words may result in full deletion of the vowel of the nucleus of a syllable, and its replacement by a syllabic consonant, such as [ŋ] in the British English *button* [ˈbʌŋ]. or by a “dark” velarized [ɫ] in the American English *bottle* [ˈbɒɫ], or by a [w] as in British English *bottle* [ˈbɒɹw]. These reductions pose difficulties for learners of English, both in terms of perception, and in terms of production (Kenworthy, 1989: 28; Celce-Murcia et al., 1996: 108). Peter Roach, although often cited as a sceptic of isochrony following the publication of his paper on the subject (1982) states in a more recent paper on RP (2004) that:

English RHYTHM is said to be STRESS-TIMED, i.e. the intervals between stressed syllables tend to be constant and unstressed syllables are compressed to preserve the isochrony of the inter-stress intervals. While the evidence for this is not completely conclusive, it is clear that in RP there is a very marked difference between weak, unstressed syllables which in some contexts may be almost undetectable and strong syllables (stressed or unstressed) which are fully pronounced. (Roach, 2004: 243)

The reduction of stressed syllables in English is often particularly difficult for French learners, both at the level of perception, and at the level of production (Pater, 1997), since not only does French not have word stress, but prominence in French is different in many aspects, as we shall see in the next section.

#### 4.1.3. *Stress and rhythm in French*

In contrast to English, French has no lexically distinctive stress (Vaissière & Michaud, 2006). In a comparative study of French, English, German and Spanish, Delattre (1965), using a corpus of 1500 French words and 5800 English words, established the following distribution of what he referred to as “logical stress” stress in those languages:

Words of:	French				English			
1 syllable	100%				100%			
2 syllables	0%	100%			74%	26%		
3 syllables	0%	0%	100%		55%	39%	6%	
4 syllables	0%	0%	0%	100%	33%	36%	29%	2%

**Table 4.1.3. A comparison of the position of stress on French and English words (adapted from Delattre 1965: 29)**

Delattre (1938: 69) was quite clear in his description of French stress patterns at the word level: “every isolated word in French has ‘rhythmical or syntactic’ stress on the final syllable”, and all other syllables are “destressed” or “unstressed”. Most authors, in traditional prosodic studies, agree that French has a single rhythmic stress (also called: logical, objective, tonic, normal, or internal stress) which is regularly assigned to the final full syllable (i.e. not containing a schwa) of the last lexical item of a stress group. The stress group can be defined in French as a prosodic unit containing a stressed syllable preceded by a number of unstressed ones. Content words (i.e. nouns, verbs, adjectives, and adverbs) are generally stressable, whereas function words are not. (Di Cristo 1998: 4). The final syllable of the stress group is also the place where the pitch contour will be realised when the stress group is final in the intonation unit. (Di Cristo, 1998: 4). In fact, Mario Rossi (1979: 39), after conducting several experiments on the production and perception of stress, concluded that French was a language without stress, in the sense that stress and intonation in French, both by their nature and by their



function, do not constitute two distinct entities.

There is, of course, prosodic prominence in French, but French prosody is primarily structured around the *groupe de sens*, (sense-group) (Delattre, 1966b; Vaissière & Michaud, 2002). These groups are also commonly known as *groupes rythmiques* (rhythmic groups), or *groupes de souffle* (breath groups), and there are several different ways of classifying them. Many authors agree that two levels of prosodic unit may be distinguished in French. The smaller of the two may be called a “stress group” (Di Cristo 1998), “accentual phrase” (Jun & Fougeron 1995), or a “melodic phrase” (Vaissière & Michaud, 2006), and it may contain several syllables or even words. Di Cristo (1998) states that French has a “rhythmic stress”, which is “regularly assigned to the final full syllable (i.e. not containing a schwa) of the last lexical item of a stress group”. He describes a stress group as “a prosodic unit containing a stressed syllable preceded by a number of unstressed ones” (Di Cristo 1998: 4). The larger prosodic units, known as “intonative units” (Di Cristo 1998), “intonation phrases” (Jun & Fougeron 1995), or “breath groups” (Vaissière & Michel, 2006) may contain several of the smaller units. Some analyses contain additional units, such as Jun & Fougeron (2000) who added an “Intermediate Phrase”.

In terms of acoustic correlates to prominence in French, numerous authors mention the group-final syllable being marked, most notably by an increase in its duration (Benguerel 1973; Di Cristo 1998; Lacheret-Dujour & Beaugendre 1999; Jun & Fougeron 2000; Astésano 2001). The final lengthening may be, at least in part, a consequence of group-final position, and a release of breath before the pause: as Astésano (2001: 54) states, this may be the result of the presence of stress, or it may be a possible component of stress. This final syllable, usually lengthened, is generally marked with a rising F<sub>0</sub>, but when the stress group occurs at the end of an intonation unit, the stressed syllable will often (at least in declarative utterances) fall both in F<sub>0</sub> and in amplitude (Di Cristo 1998: 4). However, these modifications may also result simply from the position of the syllable at the end of the rhythmic group (Faure & Di Cristo 1973: 234). The phenomenon of final syllable lengthening is also noticeable in isolated words, and this has been held to be a sort of word stress in French (Delattre 1965; Dahan & Bernard 1996). I would contest the term “word stress” to refer to this phenomenon in French, mainly because it has no distinctive value: it may be argued that it has some phonological value, in that it aids segmentation of speech into prosodic groups, but it is also largely phonetic in nature, as final lengthening is often simply a result of a little extra air escaping from the lungs at the end of an utterance before drawing breath again.

The rhythm of French has been defined as “syllable-timed” (Pike, 1945). Wenk and Wioland (1982) refute the fact that French is syllable-timed, and remind us that French syllables are produced and perceived in rhythmic groups, as in English, but that the French language usually signals the end of a

group with a lengthening of what is perceived as the final syllable of each group. The framing of the debate, they believe, is due to reasoning based on perception:

That the significance of group-final syllable lengthening should have been so widely overlooked may well be the consequence of a biased conceptual framework, reinforced by perceptual habits particularly attuned to leader-timed (English) rhythm. The inferior perceptual salience of temporal variations seems to have further contributed to the slighting of that most dependable manifestation of French accent: the additional duration of group final syllables. (Wenk & Wioland, 1982: 214)

The initial syllable of each group is usually also marked in French, albeit slightly less. After studying recordings of French from 1945 to 1975, Fónagy (1980) noted this phenomenon and referred to it as an “accentual arc” or “stress arc” (*arc accentuel*). According to Vaissière (2008), this prosodic form, which has its roots in the emotive broadcasting of the 1940s, is losing its original stylistic value as it becomes more usual in general conversational French.

The problems with stress encountered by French learners when speaking English have been investigated by several authors. Céline Horgues (2013) conducted a series of perception tests using English L1 listeners to judge the stress placement in recordings made by L1 French learners compared to the same stimuli recorded by English L1 speakers. Her results confirmed the difficulties French learners have with marking English stress. She concluded:

The results of the perception test investigating perceived lexical prominence confirmed that most prosodic contexts that were predicted to be challenging for French learners of L2 English indeed caused substantial instability in the perception of lexical prominence by native English listeners. (Horgues, 2013: 47)

Many of these production difficulties stem from the incorrect use of the four acoustic cues, the relative importance of which will be discussed in the next section.

#### *4.1.4. The relative weight of cues to stress in English and French*

Pater (1997) gives an account of how French learners of English have difficulties marking stress, and in an earlier paper (Pater, 1993) he gives a detailed account of *where* stress placement errors occur, but does not account for *how* stress is marked. Of the four acoustic cues and their perceptual cues used to mark prosodic features such as stress, we may ask ourselves whether they each contribute to the same degree in English and in French. In the previous section, we saw that duration is an important cue for marking the group-final syllable in French, and that F0 and amplitude also play a role. Before Bolinger’s pitch accent theory, it was generally assumed that the most important cue to stress in English was intensity (Bloomfield, 1933: 110) or “force” (Jones, 1918: V). In his preface to *An Outline of English Phonetics*, Jones did however predict that the evolution of technology would improve our knowledge of the relative weight of the acoustic cues to stress:

It will in fact be found that in innumerable cases the requisite change in the direction of intonation without any increase of force whatever is sufficient to produce on the ear the effect commonly described as stress. This fact will doubtless be demonstrated by experimental methods before long. At present, however, the relations between stress and intonation have not been fully investigated, and until this has been done, there is nothing for it but to treat stress in the conventionnel manner. (Jones, 1918: V)

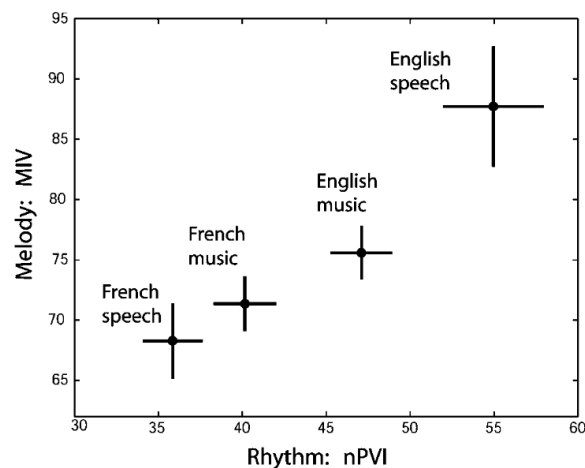
As we will see below, Jones anticipated experimental work which would follow some 40 years later. Regarding the rhythmical structure of a language, the cue which is of particular interest is duration. French, as we have seen, lengthens the last syllable of a word or a prosodic group, but the tendency to reduce vowels in English leads to much more variability in syllable duration in English than in French. One way of measuring the rhythmical structure of a language is the normalised vocalic Pairwise Variability Index (PVI) suggested by Grabe and Low (2002). This index measures the relative difference in duration of successive vocalic segments, and is normalised for local rate variations. A low PVI value shows less variation in vowel duration, and as such indicates a more syllable-timed language. Stress-timed languages on the other hand typically demonstrate shorter unstressed vowels alternating with longer vowels, resulting in a higher PVI. Grabe and Low (2002) report PVIs of 57.2 for British English and only 43.5 for French.

Language	PVI
Mandarin	27.0
Spanish	29.7
Luxembourg	37.7
Japanese	40.9
French	43.5
Polish	46.6
Rumanian	46.9
Welsh	48.2
Singapore English	52.3
Malay	53.6
British English	57.2
German	59.7
Dutch	65.5
Thai	65.8

**Table 4.1.4a. Rhythm values for 14 languages using the Pairwise Variability Index (PVI) (adapted from Grabe & Low, 2002)**

The PVI figures given by Grabe and Low in Figure 4.1.4a show that duration is much more variable in British English than in French, due to the phenomenon of vowel reduction. A more detailed comparison of the prosody of English and French is given by Patel (2006), who compared the rhythm of the speech of these two languages, to classical music composed by French and English composers. The study uses a sample of 20 read sentences in English and French, and a selection of classical music from the turn of the twentieth century. Patel used F0-based pitch contours and algorithms in Praat to calculate melodic peaks, and Grabe and Low's Pairwise Variability Index to compare syllable durations (Patel,

2006: 3034). The coefficient of variation (CV) was used in several calculations, and is defined as the standard deviation divided by the mean for duration values of syllables (Patel, 2006: 3039). In order to compare both speech and music, Patel used melodic interval variability (MIV), scaled to 100 times the CV value, to put the values in the same general range. This process allows the comparison of “rhythm-melody (RM) space” for music and speech in both languages to be represented on the same graph (Patel, 2006: 3041), reproduced in Figure 4.1.4b:



**Figure 4.1.4. Rhythm-melody (RM) space for English and French speech and music (Patel, 2006: 3041)**

Patel’s measurements show a clear difference between the rhythmical structure of speech in English and French based on the duration of the syllables, and also a corresponding difference in the pitch interval variability of the music of England and France.

As we saw in chapter one, prosody is as much a perceptual phenomenon as it is a measurable acoustic reality, for example, in rhythmical grouping (Kotz et al., 2018) as evinced by perception experiments (Jakobson, Fant & Halle, 1951; Scott et al., 1985). In the 1950s and 1960s, Dennis Fry explored the perception of stress using the Haskins Pattern Playback speech synthesiser to modify the acoustic cues to stress in a series of experiments at the Haskins laboratories (Fry, 1955; 1958, 1965). In his first series of experiments, he demonstrated the relative importance for the perception of word stress in English of duration over intensity, by modifying the cues over a series of word pairs differentiated by word stress, such as *PERmit* (noun) / *perMIT* (verb). This study had three principal findings:

- (1) duration and intensity ratios are both cues for judgements of stress
- (2) the vowel segments show the major differences in duration and intensity with a shift of stress
- (3) duration ratio is a more effective cue than intensity ratio. (Fry 1955: 767-768)

In the second experiment (Fry, 1958), he was able to add F0 to a similar protocol, proving this time that F0 was a more important cue than duration for stress perception. In the last of the three experiments of this sort (Fry, 1965), Fry again used the word pair protocol, using the noun / verb pairs

for *object*, *contract*, *subject* and *digest*. By redrawing the spectrograms by hand for the Pattern Playback synthesiser, he was able to alter the formant structure, and thereby the timbre of the vowels in the word pairs perceived by the listeners. In all of his experiments, he found that the weight of the duration cue was “very considerably greater than that of the formant cue” (Fry, 1965: 308).

Bolinger, in the same year and with the same equipment as Fry, carried out fourteen tests, the results of which are published in his paper “A Theory of Pitch Accent in English” (Bolinger, 1958). The central notion of this theory was that of “pitch prominence”, which Bolinger described as follows: “a rapid and relatively wide departure from a smooth or undulating contour” (Bolinger 1958: 112). He later clarified what these movements might be: “When only one item is given pitch prominence, it is heard as accented. The pitch movement may be UP TO, DOWN TO, or DOWN FROM the accented syllable” (Bolinger 1958: 127). As for the duration index, Bolinger's explanation is simple:

A pitch obstruction requires time for its execution. When the pitch accent is embraced completely by a single syllable, the syllable is lengthened to accommodate the necessary range of pitches...Figuratively speaking, it is there IN ORDER to make room for the accent. (Bolinger, 1958: 138)

Similar studies following Fry's initial experiments include Lehiste and Peterson (1959), who were more concerned with perception. They found that amplitude had an important role to play in correcting for changes in F0 due to the nature of the vowel (“interfering factors”):

It appears that perceptual judgements of linguistically significant stress may be based on speech power, fundamental voice frequency, vowel quality and duration [...]. The data suggest that ‘correction factors’ might be applied for the amplitudes of vowels according to vowel quality. (Lehiste et Peterson 1959: 425)

Philip Lieberman (1960) also looked at the acoustic cues of verb/noun word pairs, such as *permit*:

The stressed syllable had a higher fundamental frequency than the unstressed syllable of the same utterance in 90 per cent of the cases, a higher peak envelope in 87 per cent, and a longer duration in 66 per cent. The stressed syllable compared with its unstressed counterpart in the other word of the stress pair had a higher fundamental frequency in 72 per cent of the cases, a higher peak envelope amplitude in 90 per cent, and a longer duration in 70 per cent. (Lieberman 1960: 397)

Jenkins (1961) sought to establish a hierarchy of the three perceptual cues. He found that the order of importance is first pitch, then timbre and finally loudness (Jenkins 1961: 1557).

Although the Pattern Playback synthesiser was a great technological breakthrough, of the sort that Jones had indeed anticipated (Jones, 1918: V), it was still a relatively blunt instrument, and the vowels it produced sounded rather artificial. It was for these reasons that for my Master's degree (*maîtrise*), Daniel Hirst suggested I use modern digital speech synthesis technology to carry out a cross-linguistic study, inspired by Fry's word pair experiments. I followed classes on experimental phonetics with Albert Di Cristo, and designed my study under the supervision of Daniel Hirst, along with their colleagues in the research unit in Aix-en-Provence. We used speech synthesis software developed by

the team in Aix-en-Provence, in what was essentially a modified replication study, in that we used L1 French speakers and English L1 speakers, each listening to stimuli in their own language and the other language. From the original word pairs (*TRANSfer/transFER* and *BOITe/boiTEUX*<sup>10</sup>), we created 50 stimuli, very gradually modifying the parameters of F0, duration and formant structure from the word with initial stress to the word with final stress. When we asked the participants to say which syllable they thought was stressed on hearing all of the stimuli, each one played twice and in random order, we found that the English listeners paid more heed to F0, and the French listeners to duration. The experience of constructing and carrying out this experiment and more importantly the results led me to understand the importance of solid experimental data and especially its relevance to my own pedagogical theory and practice in pronunciation teaching. L1 French users and L1 English users, I learned, really do not perceive stress in the same way. This experiment, and everything I learned during my phonetics classes, and my two years teaching English phonetics to First year undergraduates at Aix-Marseille University set the scene for the subsequent 20 years of my research.

#### 4.1.5. *Focus, tonicity and intonation in English and French*

As we saw in section 4.4.3, prominence in French is usually used to mark larger prosodic groups, mainly by an increased duration of the group-final syllable, and to a lesser degree, by marking the group-initial syllable. We may consider this to be a sort of sentence stress, or tonic stress, but its main function is not to focus the listener's attention to new and important information, as does sentence stress in English, but rather to aid segmentation of speech into manageable chunks. As Wells (2003: 12) states: "French, however, does not use tonicity in the same way, and French learners typically have difficulties with English tonicity because of their negative transfer of the French system to English".

Although French does not have tonic stress in the same way as English, it does of course have recognisable intonation patterns over its prosodic units. English, as we saw in section 1.1.5, may be said to have seven canonical intonation patterns (Cruttenden, 1997: 54), eight with the addition of the high-rising terminal pattern. French has a similar number of patterns, with similar functions: Delattre (1966) enumerated the following ten, based on readings of classic works of French literature, including dialogues. He established a list of their functions (Delattre, 1966: 3) with four declarative forms, two

---

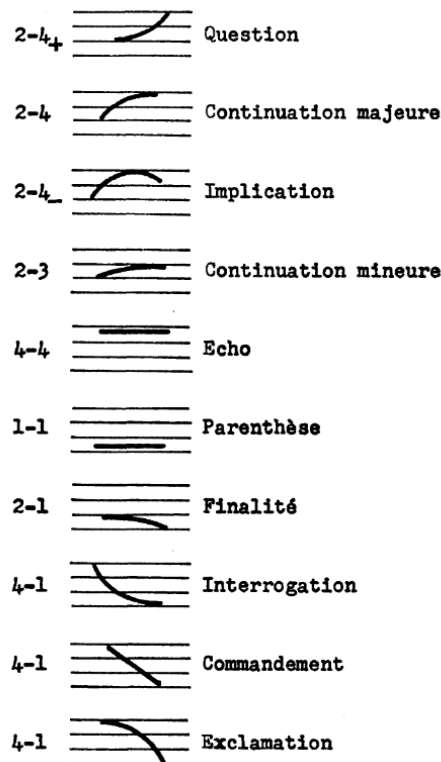
<sup>10</sup> French not being a language with lexical stress, we "cheated" in order to produce comparable stimuli: the French word *boiteux* is, at least when compared to the word *boite* (when that word is pronounced with a Provençal accent which doesn't drop the *e-muët*), comparatively more weighted on the second syllable.

interrogative forms, a parenthetical and an echo form, and an exclamation, as we can see in the following table:

finalité	}	déclaratives
continuation majeure		
continuation mineure		
implication		
commandement		
question	}	interrogatives
interrogation		
parenthèse	}	parenthétiques
écho		
exclamation		exclamative

**Table 4.1.5a. The ten canonical intonation patterns in French by function (Delattre, 1965:3)**

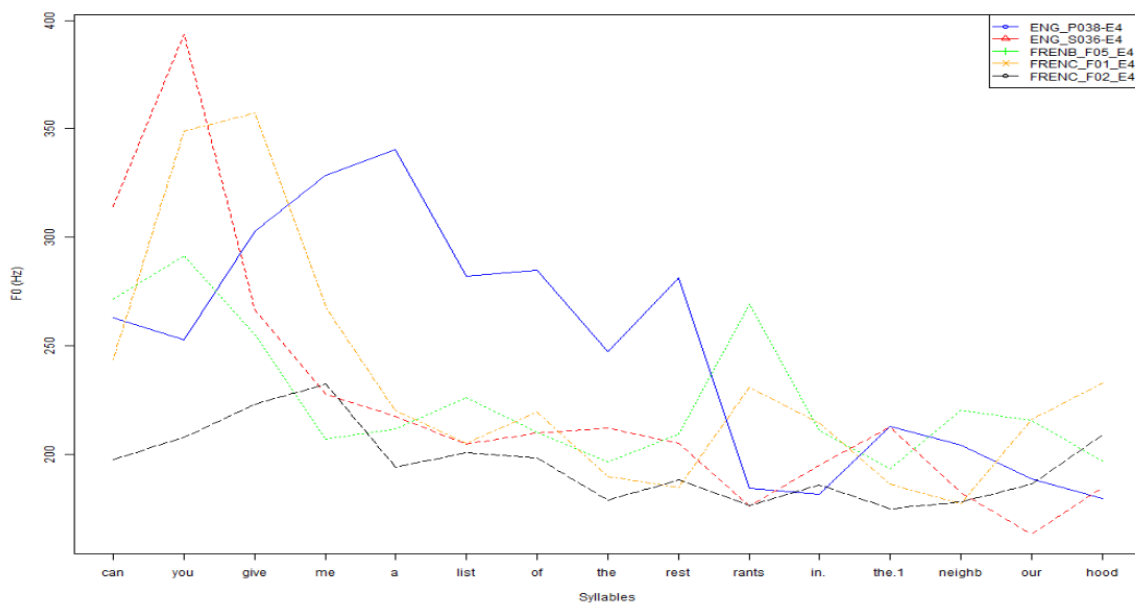
He also drew their forms, using a sort of musical stave, using four lines for low, mid-low, mid-high and high (Delattre, 1966: 4) as follows:



**Figure 4.1.5a. The forms of the ten canonical intonation patterns in French (Delattre, 1965:4)**

Based on the similarities of Delattre's 10 canonical forms and the canonical forms of English, we may assume that the intonation of the two languages is similar, and in some respects, this is true, as Hirst

and Di Cristo state (1984: 567). However, on an acoustic level, there are two important differences in the way intonation is marked in these languages: the first has to do with the range of intonation, and the second, to do with the form of its contours. Firstly, the intonational range of English is greater than for French: the peaks are often higher (although, there is, of course, much individual variation). In a study conducted by Sophie Herment and her colleagues (2014) on the intonation patterns of yes/no questions realised in English by French learners, the differences between L1 English speakers and L1 French speakers are obvious. Fig 4.1.5c shows the intonation contours for a yes/no question realised by two L1 English users, two proficient learners (C1 on the CEFR scales) and one B1 learner:



**Figure 4.1.5b. Comparison between two L1 English users, two C level learners and one B level learner (Herment et al., 2014: 4)**

The two L1 English participants (ENG\_P038-E4 and ENG\_S036-E4, the blue line and the higher of the two red lines respectively) have different intonation curves, but both start with a high rise-fall, with the falling part of the curve being fairly smooth and gradual over the whole utterance. The intonation curves of the French learners all fall quite quickly and remain below 250 Hz for most of the utterance. Even the C1 learner, FRENC\_F1\_E4, in yellow), who does manage an initial rise-fall, plateaus on *give*, indicating the syllabic realisation of this word.

Secondly, French is a language without moveable stress, and it has a more regular syllable length than English, as we have seen. This results in contours being less smooth, with more step-ups and step-downs being used between syllables, and even at the syllable level the contours are not as smooth as in English. Delattre (1963) maintains that speakers tend to impose their native intonation patterns on their second language. The following two diagrams (Figure 4.1.5d) represent the intonation patterns of an American English utterance, followed by a French utterance of a similar length and nature. The



intonation patterns are similar, but the English patterns are smoother, with curves over several syllables, whereas the French patterns tend to be less smooth, and intonation sounds “steppy”:

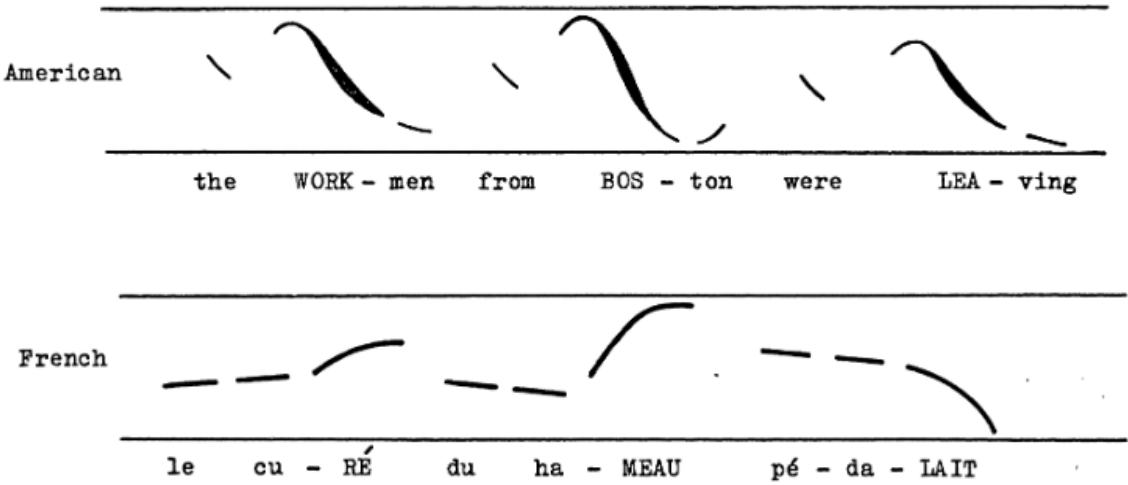


Figure 4.1.5c. Examples of English and French intonation curves (Delattre, 1963: 196)

We note that Delattre has drawn the lines by hand, and there seems to be an attempt to represent figuratively the differences he perceives in the intonation contours in these two languages, with the French curves being drawn with simple lines, and the patterns for American English being more curved, and tapering at the ends. Delattre describes his impression of American English intonation as follows: “constant recurrence of falling glides that characterizes American Intonation for the ear of foreigners” (Delattre, 1963: 197). In the next examples (Figure 4.1.5e) Delattre illustrates how the intonation pattern we saw above may be carried over into English:

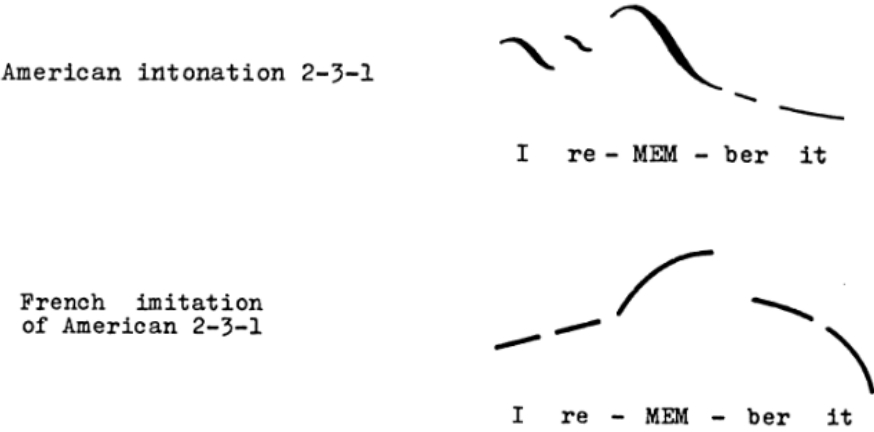


Figure 4.1.5d. Examples of how an L1 English user and an L1 French user would pronounce the same sentence (adapted from Delattre, 1963: 194)

The above example (Figure 4.1.5d) shows the “steppy” nature of French intonation carried over into English by the L1 French user, or what Herment and her colleagues (Herment et al., 2014: 5) refer to as “isosyllabic realizations, typical of French rhythm”.

In my own research, for example when calibrating the “prosody descriptors” (Frost & O’Donnell, 2018), I have noticed that although L1 French users generally master stress placement at A1-A2 level for lexical stress, it is only at B2 – C1 level that nuclear stress is systematically marked. Furthermore, across all levels except for the most proficient users, unstressed syllables are not sufficiently reduced, resulting in a perception of “steppy” rather than smooth intonation contours. This impression of a “steppiness” in French intonation that a listener has, and that Delattre represented graphically in his diagrams is compounded by more abrupt and short contours, often over one or two syllables at the end of prosodic units, rather than gradually over syllables preceding or following the nucleus. The main finding from my research regarding focus, and marking prominence generally, is that very few French learners use all four cues of stress correctly when speaking, even at the advanced C1 – C2 levels. French learners usually know *where* to place stress, but do not how to *physically* mark the prominent syllable – so it is less of a cognitive issue, and more of a physical problem. It is one thing to know that stress exists, another to know which syllable should receive it, but actually being able to physically produce the prominence, with the appropriate balance of all four acoustic correlates of prominence which provide the effect of stress on the ear of the listener is a challenge for most French speakers of English. This is one of the reasons I have adopted an embodied metacognitive approach to my teaching, as I will explain in chapter seven.

## 4.2. Vowels

The English and French vowel systems are quite different. There are fewer vowels in French than in English, and they are essentially all simple vowels, except for the four nasalised vowels /ã/, /õ/, /ɛ̃/ and /œ̃/. English vowels are traditionally grouped into short vowels, long vowels and diphthongs (Roach, 2004: 241). Vowels may also be considered to be tense, or tensing (the long monophthongs and the diphthongs), or lax (the short vowels). The tense/lax vowel alternations in English are explained by a complex set of patterns laid out comprehensively by Noam Chomsky and Morris Halle (1968). So, in addition to establishing whether a vowel is stressed, or unstressed, learners of English must also know whether to use a tensing or a lax vowel: if a vowel is stressed, it will have a full value (either tensing or lax), and if unstressed, it is likely to reduce. In the same way that marking stress is both a cognitive and a physical problem for learners as discussed previously in this chapter, so producing the correct vowel is doubly problematical. French learners generally know how to read and write words in English: the

two languages both use the Roman alphabet, and many words are transparent. However, *knowing* for example that the word “develop” is written with the letter -e- in the first two syllables, and an -o- in the last syllable, is very different from being able to physically pronounce the three vowels as /ɪ e ə /, and pronouncing the final two syllables of the word “comfortable” so that they do not sound like the word “table”. I will begin by presenting the French vowel system, and then the English vowel system.

4.2.1. French vowels

The 11 canonical oral vowels of French are shown in fig 4.2 in Figure 4.2.1. Cécile Fougeron and Caroline Smith (1993) base this representation on Parisian French, but note that varieties of French have almost identical inventories; the main differences are to be found in the maintenance or loss of certain contrasts (Fougeron & Smith, 1993: 1):

i	si	si	‘if’
e	se	ses	‘his, hers’ (pl)
ɛ	sɛ	sait	‘knows’
	sɛʁ	serre	‘tightens’
y	sy	su	‘known’
ø	sø	ceux	‘these’
œ	sœʁ	soeur	‘sister’
ə	sə	ce	‘this’
a	sa	sa	‘his, hers’ (f)
u	su	sous	‘under’
o	so	sot	‘silly’
ɔ	sɔʁ	sort	‘fate’

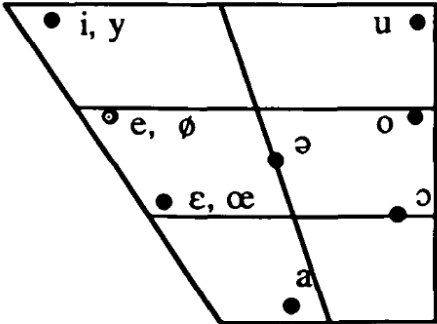


Figure 4.2.1. French oral vowels (Fougeron & Smith, 1993: 73)

Fougeron and Smith (1993: 73) also mention that some speakers differentiate between two low vowels, more central /a/ and a more back /ɑ/. They also note that although the French schwa is also a central vowel as in English, it has some rounding. The front closed, mid-closed and mid-open vowels are all differentiated by rounding, but otherwise, each pair of vowels is realised in the same place; this means that the 11 French oral vowels are realised in only eight positions. Regarding French vowels, I would simply add that the distinctions between the mid-open and mid-closed vowels are less present in the speech of younger L1 French speakers, and it may be argued that this further simplification of the French vowel system might make it even harder for French learners to realise English vowels.

#### 4.2.2. English vowels

The two varieties of English which I will deal with here are Standard Southern British English (SSBE), sometimes still referred to as Received Pronunciation (RP), or BBC accent (Roach 2004), and General American (GA), also known as Network English, or North American English (NAE) (Celce-Murcia et al., 1996). This does not represent my views on models for teaching, as we will see part two.

The short and long monophthongs in RP/SSBE can be represented as follows:

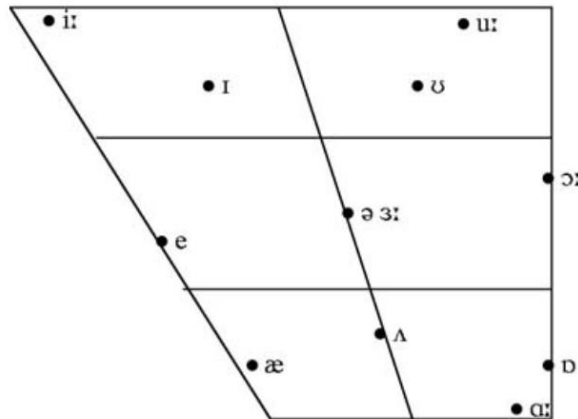


Figure 4.2.2a. Received Pronunciation: short and long monophthongs, or “pure vowels” (Roach, 2004: 242)

In addition to the 12 simple vowels shown above for RP, English also has eight diphthongs (five closing diphthongs, and three centering diphthongs), which may be represented as follows:

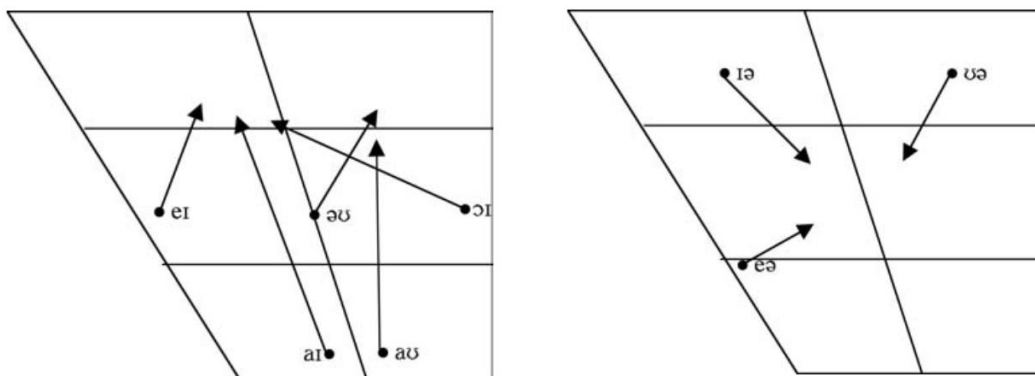


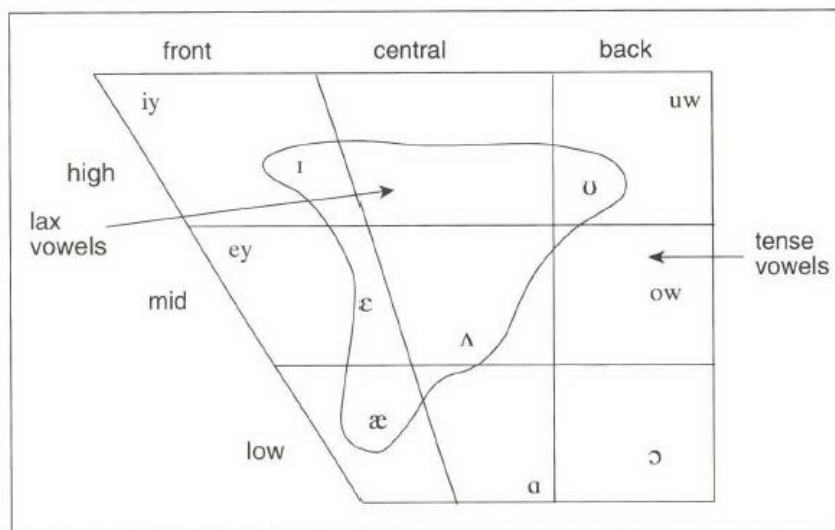
Figure 4.2.2b. Received Pronunciation: closing diphthongs and centring diphthongs (Roach, 2004: 242)

In French there are neither tensing vowels, nor diphthongs, so these vowels are often problematic for learners whose L1 or other learned languages do not contain such vowels, such as French. This problem is compounded by the fact that a stressed vowel may be either a tensing vowel /eɪ ɑ: i: aɪ əʊ (j)u:/, or a lax vowel /æ e ɪ ɒ ʌ ʊ/ (section 4.1.2). The patterns of distribution of the tensing and lax vowels was most comprehensively laid out by Chomsky and Halle (1968), and has been reworked for French

learners many times since, (Greven, 1972; Guierre, 1984; 1987; Ginésy, 1989; 1995; Lilly & Viel, 1998a; 1998b). Optimality theory and functional load theory would suggest that these vowels are all important, in the sense that they constitute the stressed syllables; the markedness differential hypothesis suggests that the vowels and the vowel oppositions which are dissimilar to French will prove more difficult. The lax / tensing opposition itself is particularly important and difficult for French learners, as the French vowel system does not have this opposition. The theories and models which I reviewed in the last chapter - together with teacher experience and an analysis of learners' needs - will enable teachers to gauge which vowels need to be worked on in any given situation. The individual vowels which are often confused include /i ɪ/, /æ ʌ ɑ: ɒ/, and /əʊ əʊ/.

Perhaps the most important vowel for learners to be able to realise and recognise is schwa, which is by far the most common reduced vowel (Flemming, 2009; Silverman, 2011). The other two lax vowels which are often a result of reductions, /ɪ/ and /ʊ/ should not be neglected either. According to Fry's study (1947), schwa is the most common of all English phonemes, accounting for 10.84 per cent of all phonemes in colloquial RP, or 27.39 of all vowels. M. Ardussi Mines and colleagues (1978) found that schwa accounted for 12.99 per cent of all phonemes in their study of American English.

The following diagram (Figure 4.2.2b) represents the trapezium of North American English<sup>11</sup> vowels (Celce-Murcia et al, 1996):



**Figure 4.2.2c. Tense versus lax vowels in North American English (Celce-Murcia et al., 1996: 96)**

<sup>11</sup> Celce-Murcia and her colleagues use the terms North American English and British English, rather than GA or RP / SSBE (Celce-Murcia et al, 1996: 35): for our purposes, these terms are synonymous.

Marianne Celce-Murcia, Donna Brinton and Janet Goodwin present the vowels of North American English with several pedagogically motivated choices, some of which I believe to be a very useful contribution to teaching resources, but not all of which work for British English. Long front and back vowels as /i:/ and /u:/ respectively, which shows graphically the tensing nature of these vowels, and introduces the linking /j/ and /w/. These approximants are often problematic for learners of English, and they help to speed up the tempo over unstressed syllables, thus contributing to the rhythmical structure of English. Schwa does not figure on this trapezium, as it is considered so important that it is mentioned apart, in a section on reduced vowels (Celce-Mucia et al., 1996: 108-110). Finally, Figure 4.2.2b contains what are generally considered in SSBE to be “pure” vowels and diphthongs in the same trapezium, leaving only three” phonemic diphthongs” /ay/, /aw/ and /ɔy/ - this would of course not work for SSBE.

Finally, unlike SSBE, in GA, is a rhotic variety: when the letter -r- appears after a vowel orthographically, this is reflected in the pronunciation of the preceding vowel by retroflexion, a curving back of the tongue. Table 4.2.2c represents this /r/-colouring of vowels in GA:

Uncolored vowel		/r/-colored vowel	
/ɪ/	lid	/ɪr/	leered
/ɛ/	fed	/ɛr/	fared
/ʊ/	should	/ʊr/	assured
/ʌ/	hut	/ʌ <sup>r</sup> /	hurt
/ɑ/	pot	/ɑr/	part
/ɔ/	caught	/ɔr/	court
/ay/	tie	/ayr/	tire
/aw/	ow!	/awr/	our

**Table 4.2.2d. /r/ coloured vowels in North American English (Celce-Mucia et al., 1996:105)**

There are other differences between SSBE and GA vowels, but for the purposes of this report, the main differences which have been presented here will suffice.

### 4.3. Consonants, approximants and allophones

I do not consider it useful here to list all of the consonants in French and in English, so I will concentrate on those consonants and the features of consonants which are useful for reasons of intelligibility, or which are relevant for prosody-related phenomena such as reduction, deletion, linking and assimilation in English. These are often what Kenworthy (1989: 123) refers to as “high priority” sounds.

It is therefore important to teach the realisation of a selection of consonants (though not *all* consonants) and this is for reasons linked to both perception and production. Phonemes, and

particularly consonants, are not just important for auditory reasons, but they also provide important visual cues. Studies have shown hyper-articulation of the syllable-initial consonants of stressed syllables can provide a useful physical aid to auditory cues (Swerts & Kramer, 2008; Rapin & Ménard, 2019). If learners can realise sounds correctly, this will provide important visible cues to listeners, and conversely, the correct realisation of consonants will aid perception and comprehension of English.

Firstly, several consonants exist in English, but not in French. This mainly concerns /θ, ð/ and /h/. It may be useful to insist on a correct dental fricative /θ/ or /ð/ for pronouncing the letters *-th-* in English if learners intend to follow a high-stakes career, such as English teacher or actor, where authentic pronunciation would be necessary. However, for the majority of learners of English, a number of allophones, even /s z/ may well be sufficient. After all, Irish speakers of English realise inter-dental fricatives as alveolar plosives (Hickey, 2004) and *-th-* in parts of London and elsewhere is often pronounced as /f/ or /v/ (Kerswill, 2006). The consonant /h/ is not difficult for French learners, but it is often dropped where it should be realised, or realised where it is present orthographically, but not realised. A dropped /h/ is not usually an impediment to intelligibility, and British English speakers often drop the /h/, even on stressed initial syllables. The nasal /ŋ/ exists in French in words borrowed from English, as do the affricates /tʃ/ and /dʒ/, the constituent sounds of which exist in French, and so they are not too problematical, either for intelligibility or for comprehension.

Secondly, there are several phonemes (consonants and approximants) which are realised differently in English and French, or which have allophones in English, some of which are different from the equivalent consonant in French. The alveolar approximant /r/, with its allophone, the alveolar tap, [ɾ], does not exist in French, and is problematical for two reasons, one phonetic, and one distributional. Phonetically, the /r/ in English is a double approximant, with the alveolar approximation accompanied by an approximation between the upper teeth and the projected lower lip. Another double articulation in English is /w/, which in French involves only the bilabial approximation, without the velar approximation so characteristic of this sound in English. In addition, the French /w/ sound also involves less projection of the lips, and so there is a different formant structure to the phoneme, and different visual cues, as with the English /r/. The semi-vowel /j/ exists in French and is realised in the same way in both languages, but it is often used for linking in English, so it can be useful to work on the three semi-vowels /w/, /r/ and /j/ for that reason. The lateral approximant /l/ is always an alveolar approximant in French, whereas in American English, it is a velarized “dark” [ɫ] in all positions, and in British English, it is velarized in syllable-final position. This is particularly worth mentioning because the velarized [ɫ] can be used a syllabic consonant in unstressed syllables, for example in the word *bottle*, /bɒtɫ/. Another allophone worth mentioning is the glottal plosive, or glottal stop. Increasingly

frequent as an allophone in all varieties of British English (Moosmüller, 1989), and even in American English, this allophone of /t/ can replace the voiceless alveolar plosive /t/ in medial or final position.

Thirdly, there are several phonemes which are almost identical in English and French, for example /p t k/ all exist in French, but they are usually aspirated in English. This is normally not problematic for intelligibility, but it may be in cases of geminate resyllabification, for example. In the following examples, the initial /p/ of the stressed syllable is aspirated, whereas a /p/ following the phoneme /s/ is not aspirated: *John's pits* versus *John spits*, or *Spielberg's prints* versus *Spielberg sprints*. This is an example of connected speech, and various connected speech phenomena may also be considered to be closely related to prosody, in that they allow for a faster speech where necessary, which makes it easier to respect the metrical constraints of English.

Connected speech phenomena are particularly important for learners at the perceptual level, because they aid comprehension (Cauldwell, 2013). These phenomena include linking, (especially with /w/, /r/ and /j/), deletion (especially of /t/ and /d/) assimilation (e.g. /tj/ in British English /ˌwɒtʃə'wɒnt/ for *what do you want?*, etc.) (Celce-Murcia et al., 1996: 158-165). Consonants can also be affected by pauses, and by the separation of speech into tone units, as these can also impact connected speech phenomena as we saw above, L2 users pause more frequently and pause in places where L1 users often don't (Hilton, 2008; Volín, 2019), so liaisons and other connected speech phenomena may be missed, making for less fluent speech, and potentially impeding comprehension.



## Conclusion (part one)

As I mentioned in the general introduction, the question which drives most of my research and a large part of my teaching and teacher training is “why do so many French people have difficulty understanding English when it is spoken by native speakers?”. By working on pronunciation, I am able to address this question, while also aiming to improve learners’ oral production skills, but my prime motivation for focusing on pronunciation is because of the relationship between perception and production, which I have approached from various angles in this part. I have explored prosody from the perspective of first language and LX acquisition, and presented several models of perception, production and memory. As the discussion about isochrony highlighted, certain prosodic features, such as rhythm, result not only from the acoustic signal, but also from the way we listen to language. In this part, I have drawn on the fields of biology, psychology and linguistics to better understand the nature and roles of prosody, from both a physical and cognitive standpoint. We have also seen that perception is an active process, and that when listening to speech, or when seeing physical cues to language such as gestures, expressions, or simply the articulation of sounds, the motor areas of the brain are triggered, and we will examine the pedagogical implications for this at the end of the next part. Chapter one concentrated on definitions of the different components of prosody and their behaviour, and in chapter two, we looked at prosody through the lens of first language acquisition. The theories and models relating to LX learning discussed in chapter three set the scene for the contrastive phonological analysis of English and French in the last chapter of this part. We have seen that prosody is the most physical and iconic part of spoken human language, that it is one of the first part of language that we develop as individuals, and that it was present at the very beginnings of human language. We have also seen the essential role that language plays in defining the identities of individuals and their place in language communities. Finally, we have seen the differences between English and French prosody, and some of the challenges this represents for language learners and teachers. In the next part of this report, we will turn towards issues more directly concerned with learning English as a foreign, second or other language, both inside and outside the classroom.

## PART TWO

### Prosody in LX teaching and research

#### Introduction

Whereas part one of this report focused on the nature of prosody, and on its roles in the reception and production of language in L1 and LX contexts, in this part, we will concern ourselves with the pedagogical aspects of the issues which this report focuses on, namely how prosody fits into the teaching and learning of English, especially in the context of France. As I mentioned in the general introduction and in the conclusion to part one, at the beginning of my career in France, my work was motivated by the problems many French learners have in understanding spoken English. My teaching and research is therefore largely concerned with problems relating to the reception of English: I work on pronunciation in order to improve communication, both in terms of reception and production.

In the twenty-five years I have been teaching and researching in France, there have been many changes regarding the status of English as a language in the world, and in terms of the way learners engage with English here in France. The “*Loi Toubon*”<sup>12</sup> of 1994, and the various pronouncements of the *Académie Française* on what, how and where French ought to be used in France now have less weight than ever compared to the many video and audio content streaming services, social media platforms, the international music industry, and the gaming industry, all of whom produce easily accessible content in English. In chapter five, we will therefore begin by exploring the place on English in the world, and in France, and some of the various policies and initiatives around languages in primary, secondary and tertiary education in France.

After presenting the context, in chapter six we will examine some of the research around the role of prosody in accent, intelligibility and comprehensibility studies in LX pronunciation teaching. As we will see, these constructs are not always well-defined, and this is problematical on several levels. At the end of chapter six, I will suggest how these definitions may be improved, and I propose a model of

---

<sup>12</sup> La Loi n°94-665 du 4 août 1994 relative à l'emploi de la langue française  
<https://www.legifrance.gouv.fr/loda/id/LEGITEXT000005616341>

language perception and production which I hope will clear up some of the ambiguity around these two terms.

Chapter seven is devoted to what and how prosody can be integrated into English teaching programmes. First, I explore the question of models (in the sense of the variety of English various stakeholders consider useful for teaching English), and questions such as the “nativeness principle” (Levis, 2020), which pits native speaker-like proficiency versus intelligibility as the ultimate goals of pronunciation teaching. In terms of language teaching programmes, I consider assessment to be essential: for institutions, for teachers and most importantly, for learners, it is necessary to understand learners’ strengths and weaknesses, in order to set learning objectives. In this final chapter, I will present an instrument for measuring pronunciation which draws on what existing research and my own research has shown us about the role of various prosodic and segmental features both in terms of reception and in terms of their contribution to intelligibility. This is followed by some suggestions for both content and activities for a prosody-centered approach to teaching English pronunciation to French learners, using embodied activities and metacognition, in order to develop both reception and production skills. Finally, after examining some of the ethical questions that pronunciation teaching raises, I outline some future directions for research, teaching and teacher training.

## CHAPTER FIVE

### English and English phonology in the French educational context

Johanne Paradis describes factors which influence language acquisition in childhood in terms of internal and external variables (Paradis, 2007). Having examined some of the internal factors in the previous chapter, let us begin this part by examining some of the external, or societal factors and how they can be incorporated into teaching and researching pronunciation.

In the context of LX learning, especially when it comes to English, very few learners are exposed to the target language in the classroom alone. Ecological models, such as Bronfenbrenner's ecological systems theory, with its five systems model (Bronfenbrenner, 1974; 1977; 1994), illustrate the various contexts in which development takes place for a child, but much of this theory also applies to older learners. Typically, older learners of an LX in France have relatively few hours in contact with a language in the classroom, but they may have to use the target language frequently in their professional or personal lives. This is particularly true of English, as we will see in section 5.1. If as teachers and researchers we remember that this is the case, it can help us when we negotiate goals and strategies together with learners, and with other stakeholders as necessary, so they can continue to learn outside the classroom. Setting goals for pronunciation teaching therefore means thinking about the needs and requirements of the language communities where learners will be using English: as Jonathan Leather (2008) points out: "In terms of acquirers' implicit goals, socialization can be seen as progressive alignment of their individual speech with the phonological profile of a community." (Leather, 2008: 51). This is as true of EFL learners as it is of ESOL learners in an English-speaking country, or of individuals in multilingual societies. In Leather's model of an ecological approach to phonological acquisition, he situates the learner not only at the heart of an ecology of systems, as does Bronfenbrenner, but also draws on several complementary models (including optimality theory) to explain the complex set of interactions involving the individual in all of these systems – after all, one single model or theory cannot explain so many different processes in so many different contexts. In Leather's (2002) model pictured in figure 5, optimality theory is at the centre of the system, but the individual's phonological acquisition is part of a "complex adaptive system" which is situated in a social environment, where "Social knowledge mediates between the acquirer's lexicon (in the broad sense) and phonological constraints of the embodied situation" (Leather, 2002: 62).

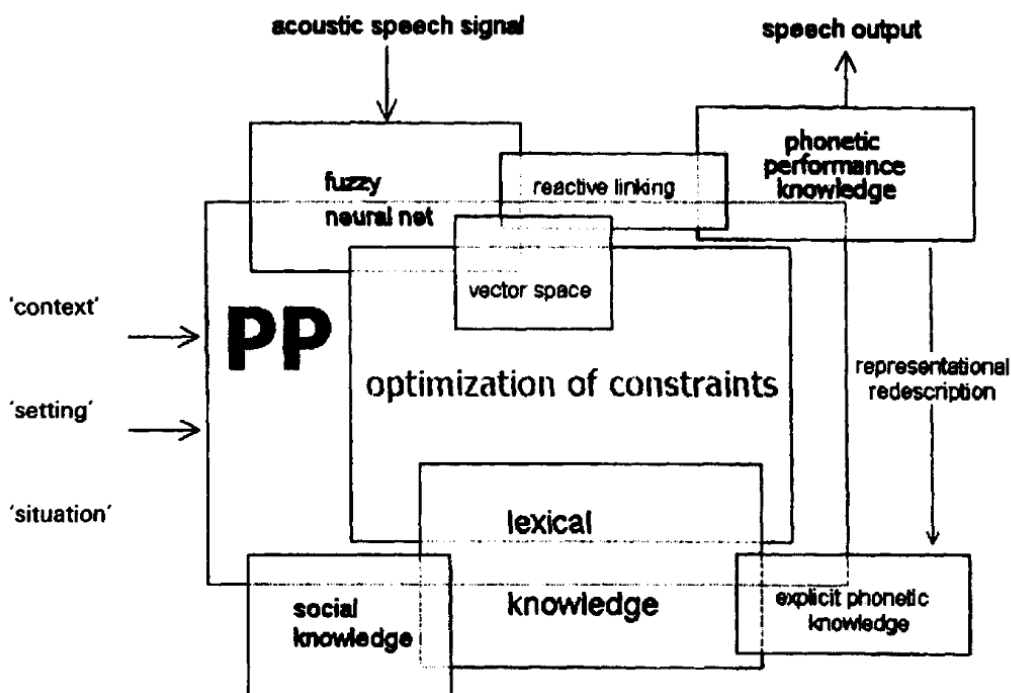


Figure 5. An outline of a syncretic model of phonological acquisition grounded in phenomenological phonology (PP) (Leather, 2002: 62)

So, if we are to accept that each learner's "ecology" is different, then we cannot teach all our learners the same set of knowledge and skills, we must help them to learn according to the set of circumstances in which they find themselves. In terms of learning pronunciation, some of that knowledge and some of those skills will be common to all learners in a class (for example, to all L1 French users learning LX English), but according to their interaction with the target language, they will need to employ different strategies. In this chapter, we will examine the different factors which make up the complex learning situations of LX English learners in France today, starting with the status of English in the world.

### 5.1. World Englishes

Before focusing on English teaching in the French context, it is important to situate English in a wider context. With the level of real and virtual mobility for personal, professional and study reasons that life today involves, learners in France are more likely than ever to need English to communicate, both inside and outside France. When we add to this the contact with English through music, streamed media, and social network platforms, there has never been so much potential for contact with different varieties of English for learners of English the world over. In the 1980s, it was thought by some that with the number of varieties of English that were present in the world, that the diaspora of Englishes would inevitably end in mutually incomprehensible varieties with a century (Quirk & Widdowson,

1985). In the three-circle model, first suggested by Kachru (1985) and represented in Figure 5.1a (MacArthur, 1987: 11), these potentially mutually incomprehensible varieties form part of the outer circle, where English is in direct contact with other languages. The model describes a core of “world standard English”, which shares certain common features, a second circle of regional varieties and an ever-expanding outer circle of varieties of English on a country by country, or even region-by-region basis:

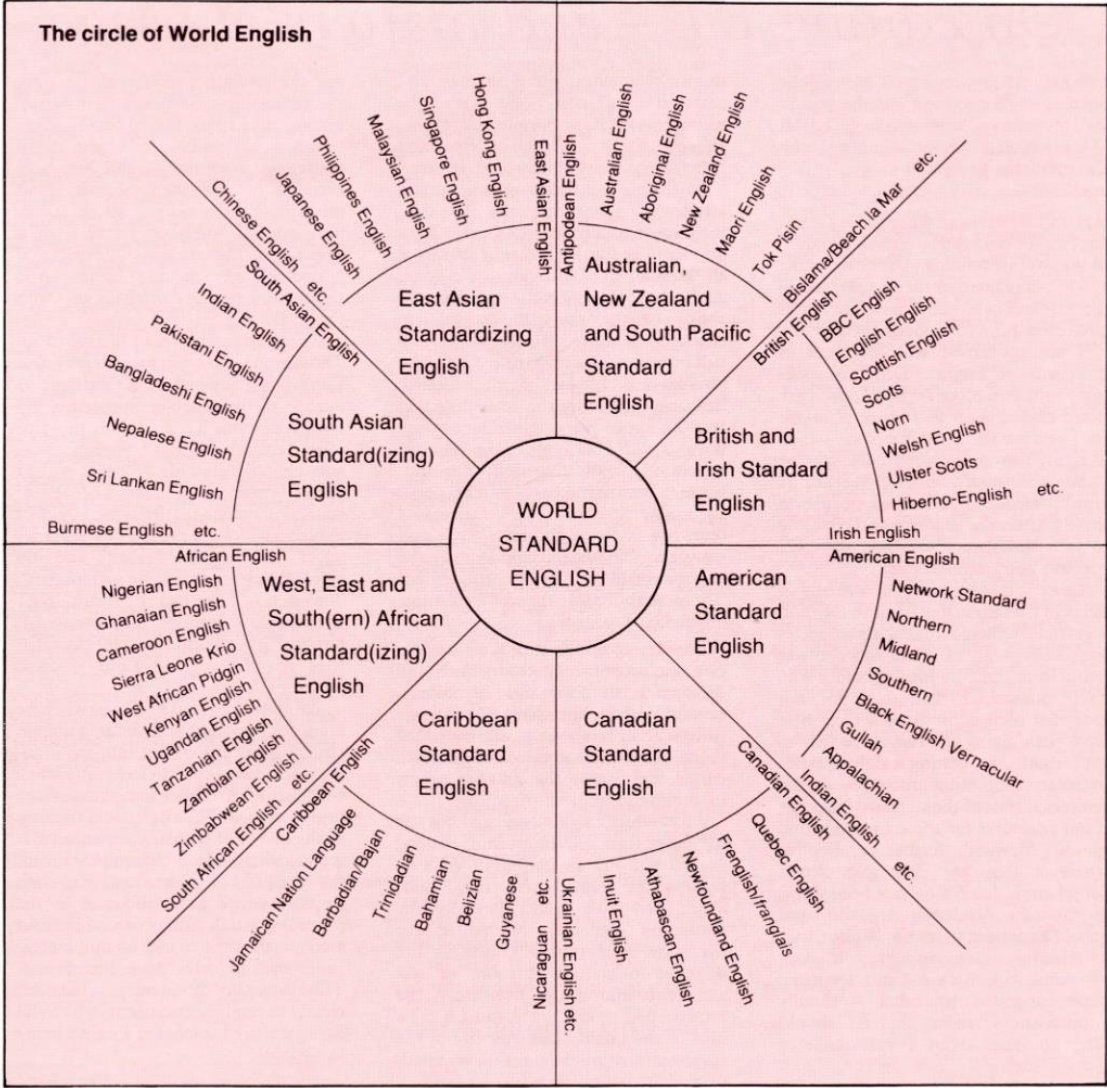


Figure 5.1. Kachru’s (1985) three-circle model of world English (MacArthur, 1987: 11)

Since the three-circle model was proposed, interconnectedness on a global scale, both physically and virtually has increased even more rapidly, and the beginning of the 21<sup>st</sup> century saw a turning point in the status of English as a world language. Using Kachru’s model, Crystal (1997: 54) estimated that at the end of the 20<sup>th</sup> century, the inner circle represented 320-380 million English users, the outer circle represented 150 – 300 million users, and the expanding circle represented 100 – 1000 million users. It

is estimated that in 2000, the number of exchanges between L1 speakers of English was outnumbered by the number of exchanges between LX speakers of English. As Christopher Brumfit (2001:116) states: “the current competent users of English number up to seven hundred million, living in every continent [...] of whom less than half are native speakers. Statistically, native speakers are in a minority for language use”. Different authors have referred to what Kachru called “World English” (1985), or to similar versions of the notion of English as a world language by various names, suggesting descriptions and definitions. Jennifer Jenkins uses the terms “English as a lingua franca” (2000; 2002; 2009; 2018). Barbara Seidlhofer prefers “International English”, whereas other authors, such as Alex Baratta (2019) refute the concept of a single core variety and prefer to talk of “World Englishes” in the plural. There are many reasons why these terms are important, not least of which is the question of language teaching programmes and which variety or varieties of English to teach.

Jenkins (2000) laid out what she called “*The Phonology of English as an International Language*” (EIL) in her eponymous book. She based her conclusions on extensive research of exchanges in English between LX users of English in many different contexts and the communication problems which often occurred. Based on this research, she made several recommendations which have proved controversial, not least of which is the central premise of her book, that English as a Lingua Franca can be distilled to a “Lingua Franca Core (LFC), comprising a set of phonological properties which can and should be taught on English-teaching programmes the world over. Following her analysis of Interlanguage talk, Jenkins mentions far more segmental features in the context of intelligibility problems than suprasegmental features, such as syllable final /n/, /m/ and /ŋ/ (2000: 33) or syllable-initial /l/ and /r/ in Japanese English interlanguage talk (2000: 34). The following table (table 5.1) proposes “native speaker” targets, which Jenkins argues against, in favour of EIL targets based on her notion of the Lingua Franca core:

	NS target	EIL target
1. The consonantal inventory	<ul style="list-style-type: none"> <li>• all sounds</li> <li>• RP non-rhotic /r/</li> <li>• GA rhotic /r/</li> <li>• RP intervocalic [t]</li> <li>• GA intervocalic [r]</li> </ul>	<ul style="list-style-type: none"> <li>• all sounds except /θ/, /ð/ and [ʔ]</li> <li>• rhotic /r/ only</li> <li>• intervocalic [t] only</li> </ul>
2. Phonetic requirements	<ul style="list-style-type: none"> <li>• rarely specified</li> </ul>	<ul style="list-style-type: none"> <li>• aspiration after /p/ /t/ /k/</li> <li>• appropriate vowel length before fortis/lenis consonants</li> </ul>
3. Consonant clusters	<ul style="list-style-type: none"> <li>• all word positions</li> </ul>	<ul style="list-style-type: none"> <li>• word initially, word medially</li> </ul>
4. Vowel quantity	<ul style="list-style-type: none"> <li>• long-short contrast</li> </ul>	<ul style="list-style-type: none"> <li>• long-short contrast</li> </ul>
5. Vowel quality	<ul style="list-style-type: none"> <li>• close to RP or GA</li> </ul>	<ul style="list-style-type: none"> <li>• L2 (consistent) regional qualities</li> </ul>
6. Weak forms	<ul style="list-style-type: none"> <li>• essential</li> </ul>	<ul style="list-style-type: none"> <li>• unhelpful to intelligibility</li> </ul>
7. Features of connected speech	<ul style="list-style-type: none"> <li>• all</li> </ul>	<ul style="list-style-type: none"> <li>• inconsequential or unhelpful</li> </ul>
8. Stress-timed rhythm	<ul style="list-style-type: none"> <li>• important</li> </ul>	<ul style="list-style-type: none"> <li>• does not exist</li> </ul>
9. Word stress	<ul style="list-style-type: none"> <li>• critical</li> </ul>	<ul style="list-style-type: none"> <li>• unteachable/can reduce flexibility</li> </ul>
10. Pitch movement	<ul style="list-style-type: none"> <li>• essential for indicating attitudes and grammar</li> </ul>	<ul style="list-style-type: none"> <li>• unteachable/incorrectly linked to NS attitudes/grammar</li> </ul>
11. Nuclear (tonic) stress	<ul style="list-style-type: none"> <li>• important</li> </ul>	<ul style="list-style-type: none"> <li>• critical</li> </ul>

**Table 5.1. EIL (English as an International language) and NS (native speaker) pronunciation targets (Jenkins, 2002: 99)**

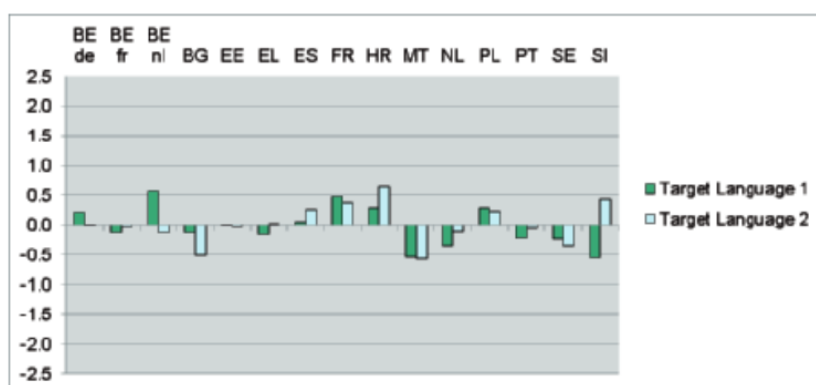
I disagree strongly with several of the points given by Jenkins to support the pronunciation targets in table 5.1, for example teaching weak forms may not be useful for intelligibility, but they must be taught for reasons of perception and access to oral comprehension (Cauldwell, 2013, Hancock, 2020; Couper, 2022a). Furthermore, Jenkins' assertion that word stress and pitch movement are "unteachable" is not only unfounded, it has been proven false by numerous studies (including my own work) as we shall see in chapter six. Jenkins proposed these ideas over 20 years ago, and as she stated (2000: 32) on the subject of the relative contribution of segmental or suprasegmental features to intelligibility that at the time of writing "no serious comprehensive investigation of the relative contribution to intelligibility of these two areas has been conducted at all; let alone within the context of ILT (interlanguage talk)". However, as we will see in chapter six, there had been some studies at the time Jenkins was writing, and there have been many since which show not only that these features are teachable, but also linking them to accentedness, intelligibility and comprehensibility. Perhaps the most positive outcome of



Jenkins' work on English as a Lingua Franca phonology (2000; 2002), is the discussion and further research it has provoked. Jenkins' work was heavily skewed in favour of East Asia, and it does not, and cannot represent all exchanges in English the world over. Seidlhofer (2001) called for more empirical research to fill these gaps, and there are now many corpora of different varieties of English which can be consulted and conclusions drawn. These corpora include the first ELF corpus, the Vienna-Oxford International Corpus of English (VOICE), followed by the corpus of English as a Lingua Franca in Academic Settings (ELFA), in Tampere, and now Helsinki University, the Asian Corpus of English (ACE) in Hong Kong (Jenkins et al., 2011), and many others since.

## 5.2. The status of English as an LX in France

According to a survey carried out by NS Opinion & Social and commissioned by the European Commission in 2012 (European Commission, 2012a), 39% of French citizens consider themselves capable of having a conversation in English (the average for all European member states being 38%, and the highest percentages being found in the Netherlands and Sweden, at 90% and 89% respectively). In the same year, European Commission also carried out the largest study of its kind ever conducted in Europe, the European Survey on Language Competences, or ELSC (European Commission, 2012b), which provides a wealth of information on the skills in various languages throughout the European Union. In the results to the teacher questionnaire, we see that teachers in France reported that they placed a relatively large emphasis on the teaching of pronunciation, as we can see in figure 5.2:



**Figure 5.2. Relative emphasis teachers put on pronunciation, ESLC (European Commission, 2012b: 204)**

The data represented from the teacher questionnaire in Figure 5.2a would suggest that English teachers in France place great store on pronunciation, but as we shall see in section 5.4, teacher training in the area of pronunciation is some way behind what these data may lead us to expect.

Regarding the demand for English in the jobs market in France, English is very much in demand. In a survey of 88 small, medium and large companies in France (Braud et al., 2016), English was reported to be “indispensable” by 51.8 % of respondents, and useful by a further 27.8 % of respondents. However, in the follow-up interviews, many professionals mentioned that the level in oral English of their employees often caused communication problems with L1 English speakers and with other English LX speakers from other countries (Braud et al., 2016: 32). It seems then that the French education is not matching the aspirations of its teachers, or of its learners. In the next section, we will explore how pronunciation is taught in the various contexts where English is taught in France.

### 5.3. Language education in France: a complex picture

In France, most civil servants - including teachers - are recruited via a competitive examination process known as a *concours*, typically composed of a written exam. This is a national process, the Ministry of Education setting a quota based on the number of teachers required which determines the pass rate. A newly qualified high school teacher will probably have to move to another part of France where there is a need for teachers, at least for the first few years of their career, and primary school teachers are generally appointed within the region of France where they trained. Once a teacher has finished their initial training, there is no system for professional development, except on an *ad hoc*, voluntary basis, generally via seminars and short courses at local universities. Teacher training has been increasingly integrated into the university system, first by the creation of the teacher training colleges in 1990, known initially as *instituts universitaires de formation des maîtres* (IUFM), then by the 2005 education reforms, when they were rebranded as *Instituts nationaux supérieurs du professorat et de l'éducation* (INSPE). Trainee teachers will have already completed a Bachelor's degree or a Master's degree in Modern Languages or Applied Modern Languages. There is no one approach or methodology which dominates language pedagogy in teacher training colleges currently: Christian Puren referred to the eclecticism of language pedagogy in France in the 1990s (Puren, 1994), however the CEFR has had a major impact on the way languages are taught in France. The actual content of the courses in French teacher training colleges varies greatly locally, and embodied teaching seems to remain marginal, as Jean-Rémi Lapaire's (2014) survey showed. Lapaire and his colleagues conducted a survey in the teacher training college at Bordeaux-Montaigne University, 140 trainee teachers responded to the questionnaire, and the findings showed that this was an area rarely dealt with in teacher training: of the 107 people who had studied linguistics, three-quarters had never studied co-verbal gestures (Lapaire, 2014a: 8). In interviews carried out by researchers at the same teacher training college,

several interviewees even stated that “a good class is a class where the pupils do not move, except to raise their hands” (Lapaire, 2014a: 7).

Since the beginning of the 21<sup>st</sup> century, there have been three sets of reforms in France related to language learning in the French education system, from three different Ministers of Education. It is not my intention to describe these projects in detail, but I would like to draw attention to some points which are relevant to the learning of oral English, and to the place of pronunciation teaching, especially to prosody, where it is mentioned in official documents. In 2000, Jack Lang, the new Minister of Education, together with Catherine Tasca, Minister for Culture and Communication, established a five-year plan which was very much structured around a desire to increase access to culture in the classroom. As part of this plan, the national inspector of schools, François Goullier was tasked with setting out the objectives for the education system at primary, secondary and tertiary level in terms of modern languages. Goullier had participated in the creation of the *Common European Framework of Reference for Languages* (Council of Europe, 2001; 2018), and was instrumental in the application of the action-oriented approach in French schools, and in setting achievement objectives for four key stages (*cycles*), covering primary and secondary education. It was therefore through legislation, in the form of a law on the school syllabi which was promulgated in April 2005, that the next Education Minister, François Fillon institutionalised the CEFR as part of the “mastery of languages”, which constitutes the “second pillar” of the foundation (*le socle commun*) of the national curriculum (Buisson-Fenet, 2014: 90). This made France the only member state of the European Union to integrate the CEFR into its legislation (Buisson-Fenet, 2014: 87). Current targets, based on the 2005 report, are that pupils should be A1 on leaving primary education (11 years old), A2 at the end of the first year of middle school (*collège*), B1 at the end of obligatory schooling, and B2 for pupils who take the *Baccalauréat* at the end of high school (Ministère de l’Éducation Nationale, de la Jeunesse et des Sports, 2006; 2007).

Jean-Michel Blanquer, Minister of Education from 2017 to 2022, commissioned a report on language teaching in French national Education. This report, entitled *Propositions pour une meilleure maîtrise des langues vivantes étrangères. Oser dire le nouveau monde*<sup>13</sup> (Taylor & Manès-Bonnisseau, 2018) describes the current state of modern language teaching in French schools and makes a number of recommendations. As concerns oral English, these include raising the level in oral English expected of pupils at the end of key stages 3 and 4, improving teacher training for future primary school teachers, and lowering the age at which children begin to learn English so they can start as young as possible. On the subject of teaching in primary schools, Taylor and Manès-Bonnisseau (2018: 65) state “In

---

<sup>13</sup> Proposals for a better mastery of foreign languages. Dare to say “new world”.

primary education, begin early, very early, as early as possible, according to researchers. Starting in primary schools, and particularly in English, work on musicality, phonology and stress in the language". Based on these findings, the Ministry has published a new action plan to improve the teaching of foreign languages throughout the period of obligatory schooling (Ministère de l'Éducation Nationale, de la Jeunesse et des Sports, 2022), which does indeed make some concrete proposals, notably for primary schools, as we shall see in the next section.

### 5.3.1. Primary education

English was introduced into primary schools by Lionel Jospin when he was Minister for Education in 1989. As part of the school reform law, he set up an experimental programme to teach foreign languages, during the last two years of primary school, which was called *l'Enseignement d'Initiation aux Langues Etrangères (EILE)* (Ducancel, 1992). In 1995, his successor François Bayrou introduced his plan, *Initiation aux Langues Vivantes (ILV)*, which recommended smaller classes, and emphasised the importance of oral training. By 2003, 96% of pupils were receiving English classes in *CM2* (the last year of primary school) but still only 56% in *CE2* (pupils aged 8-9) (Legendre, 2003:13). The situation has gradually improved over the years, but there are still very few primary school children who receive four years of English classes with a teacher who is trained in teaching English to children before they reach high school. This recruitment and training of teachers was transferred into a Master's degree in 2010, and I was part of the pilot committee who designed the content for the two-year Master's programme at Grenoble Alpes University for primary school teachers. The new Master's degree, originally called *Métiers de l'enseignement scolaire (MES)* was renamed the following year and is now known as *Métiers de l'enseignement, de l'éducation et de la formation. (MEEF)*. As the future primary school teachers were to teach languages, they were required to follow 64 hours of classes in English and in how to teach English, both of which I taught for several years, where I focused on oral work, emphasising the importance of prosody, and using songs, nursery rhymes, games, etc. The first cohort was required to achieve B2 level, attested by the French language certificate *Certificat de compétences en langues de l'enseignement supérieur (CLES)* in order to attain their teaching certificate, but this requirement was dropped in subsequent years, and is now a recommendation, not a requirement. Despite the rhetoric of successive Ministers of Education, the number of hours available for teacher training and for improving the language skills of these young teachers has been progressively reduced due to cuts in funding, and the current teacher training programme at Grenoble Alpes University contains a module of nine hours in the first year and 12 hours in the second year, including a few hours of training in how to teach English to younger learners. The disparate levels of trainee teachers and

the lack of training is commented on in the report by the national inspectorate on the implementation of the first year of the new Masters MEEF for future primary school teachers, the *parcours préparatoires au professorat des écoles (PPPE)*:

For the teachers, the problem lies mainly in the great heterogeneity of the students' level, which ranges from a very fragile level to almost perfect fluency. The objective of this first year was to reinforce the B1-B2 level. However, some students arrive with a level well below this. One teacher describes the profile of his students as follows: "As far as oral skills are concerned, one third have an A2 level, one third B1 and one third B2 to be consolidated"; some teachers go so far as to position their students at A1 level. Adaptations have been necessary. Some teachers suggested increasing the number of hours. (Hunault & Leloup, 2022)

The content of the national curriculum for primary schools as regards teaching modern languages, begins at *CP level*, when pupils are 6 years old (*cycle 2*) with oral language, and the aim of achieving A1 level for the pupils after three years:

acquiring knowledge primarily through oral expression, and exercising one's ear to the sounds of a new language are the objectives of this teaching, which must consider the age, abilities and interests of the pupils (Ministère de l'Éducation Nationale, de la Jeunesse et des Sports, 2007: 22)

A2 level is the target for the end *cycle 2*, which constitutes the last two years of primary school and the first year of secondary school. However, since training for future primary school teachers to help them achieve the aim of developing the language skills of their pupils from age eight to eleven has all but vanished from the teacher training colleges, more pragmatic solutions have had to be found to meet the ever more ambitious aims for language teaching in schools. Although not a satisfactory substitute for actual teacher training, we may cite Chantal Manès-Bonnisseau's 42-page guide aimed at primary school teachers (Manès-Bonnisseau, 2019). This extremely accessible guide refers to research, provides advice, and suggests resources for teaching pronunciation to younger learners. After briefly explaining the concept of linguistic transfer, Manès-Bonnisseau stresses the importance of phonology in teaching younger learners. She mentions how L1 prosody is learned in the womb, and how prosodic cues aid segmentation in perception, and then makes the following recommendations:

It is relevant to focus first on working on the music of the language, before the sounds themselves. In English, the aim is to train pupils to perceive and produce intonational and rhythmic patterns, i.e. the alternation between strong and weak beats. For this work on rhythm, it is then interesting to engage the child's body so that it can express itself fully. (Manès-Bonnisseau, 2019: 11)

As I mentioned above, most teachers have no real training in how to do this, so there have been many attempts to develop tools which could provide some concrete help to primary school teachers. I will simply present two solutions which have been developed with this aim, both of which are mentioned in this guide. Firstly, Anne-Marie Voise developed *Roxy and Me*, a package consisting of a glove puppet, a comic-book style text book and a CD recorded by a professional story-teller from England (Voise, 2018). In this way, with a minimum of actual speaking on the part of the teacher, the pupils may be

exposed to authentic-sounding English, albeit through the medium of recorded speech and a glove puppet.

More recently, a vocal assistant named Captain Kelly has been developed along with a suite of online resources aimed at younger learners (Ministère de l'Éducation Nationale, de la Jeunesse et des Sports, 2022). The variety of English which has been chosen is British English, and although there are some activities aimed at phonology, most of the resources are aimed at vocabulary and grammar:

In cycle 2, teachers can complement their teaching with the Captain Kelly voice assistant, a digital resource that helps to work on students' lexical and syntactic knowledge and to improve their listening comprehension and pronunciation in English. (Ministère de l'Éducation Nationale, de la Jeunesse et des Sports, 2022)

The ambitious aims laid out in the ministerial document cited above seem to be at odds with the reality which is the lived experience of many primary school teachers, that of linguistic insecurity due to their own lack of language skills and pedagogical skills in the area of language teaching (Roussi & Cherkaoui Messin, 2011; Behra, 2019).

### 5.3.2. *Secondary education*

Before we discuss English teaching in schools, relevant elements of the national curriculum, and some of the resources available, let us return to the recruitment process for teachers in French high schools, (bearing in mind that a proportion of these teachers also end up teaching in the tertiary sector). The *concours* for recruiting secondary school teachers, the *CAPES* and the *Agrégation*, and their content and form is decided on nationally and fixed by ministerial texts. Teachers who have passed either of these *concours* are also eligible for permanent teaching positions in the tertiary sector. The *Agrégation* is more demanding and prestigious, and teachers who pass are considered as part of an elite, earning a higher salary progressing in their careers more quickly (Miras et al., 2022). The *concours* are usually taken by students enrolled on the two-year teacher training programme (*Masters MEEF*) for English teachers, which itself has no fixed content, although recommendations are laid out in a framework document. Following the *CAPES* and the *Agrégation*, a report is written by the president of the exam board (*les rapports de jury*) outlining requirements and performances of each year's cohort of candidates, and in section 5.4, I will present some of the issues relating to phonology as outlined in these reports.

Candidates for both the *CAPES* and the *Agrégation* are tested on their formal knowledge of phonology in English. They are expected to be able to read and write the characters of the International Phonetic Alphabet, and know the forms, distribution and functions of word stress, sentence stress and

intonation patterns. For the *Agrégation*, requirements are higher, and if a candidate chooses the *Option Linguistique*, they are expected to have extensive formal knowledge of morpho-syntax and phonology. Based on an analysis of the *Rapports de jury* for the *CAPES* and the *Agrégation* for the last five years, the two varieties of English which are clearly favoured are General American (GA) and Standard Southern British English (SSBE) (Wilson, in press).

At *CAPES* level, candidates are expected to be able to recognise and explain the differences between SSBE and GA. Last year's question was on word stress, and candidates were expected to know the patterns of word stress placement, which some were able to do, but as Marena Turin-Barthier, president of the *jury du CAPES*, a "non-negligible number", were not (2021: 74-75). She also comments at length (2021: 103-104) on the inability of many candidates to place word stress correctly when speaking English, and to reduce unstressed syllables to schwa, giving many examples.

For the written part of the *Agrégation*, candidates are also expected to recognise and explain various features of the two "reference varieties" of English (SSBE and GA) using appropriate metalanguage, and they must say at the beginning of the exam which variety they plan to use (Gillisen, 2021: 35). At the oral exam stage of the *Agrégation*, candidates are now allowed to use any variety of spoken English they wish to, but their accent must remain consistent, (Torrent, 2022: 43). It is certainly true that a future English teacher, at any level of education, should be expected to have a very good level of English, however, the question of whether SSBE and GA should continue to be held as the only reference models for English teaching in France could be held to be another example of "native speakerism" (Holliday, 2006) inherent in the French national education system (Miras et al., 2022; Wilson, in press).

The three benchmarks during secondary education are A2 level at the end of the first year of high school (*cycle 3*), B1 at the end of middle school (*cycle 4*) and B2 at the end of compulsory schooling for pupils taking the *Baccalauréat* (Goullier, 2005: 38). Acting on the recommendations (Manès-Bonnisseau & Taylor, 2019) following the 2018 report a placement test was taken by all pupils at the end of middle school (*troisième*) for the first time at the end of the school year 2021-22. The test, *Ev@lang collège*, was developed by France Education International (FEI), and tests oral and written comprehension and the items are calibrated for A1 to B1+. According to the official ministerial documentation:

At the international level, language skills are at the heart of an ambitious plan to establish a European education area, which will come to fruition in 2025 with the Pisa 2025 tests of foreign language skills among 15-year olds. This assessment will focus on English, the most widely taught foreign language in the world. (Ministère de l'Éducation Nationale, de la Jeunesse et des Sports, 2022)

For reasons of practicability and costs, this test does not measure oral production, however the national curriculum for middle schools (*cycle 4*) does stress the importance of phonology, stating the following objectives:

- Become aware of the regularities of oral language.
- Become aware of phonic and phonological variations in the uses of the same language.
- Aim for fluency, intelligibility, personal language security in oral production: do not aim for “native accent”. (Ministère de l’Éducation Nationale, de la Jeunesse et des Sports, 2020: 45)

The national curriculum (*cycle 4*) mentions the importance of the correct intonation and appropriate gestures and body language in this context (Ministère de l’Éducation Nationale, de la Jeunesse et des Sports, 2020: 37)

In addition to teaching English in dedicated English classes, some primary and secondary schools offer content and language integrated learning (CLIL) or English medium instruction (EMI) courses, which are which are usually the initiative of local teachers, and which are sometimes encouraged by local education authorities. The Rectorate of Grenoble pays for a CLIL training course which is run by the *Service des langues* at Grenoble Alpes University which I helped to set up two years ago and have been teaching on ever since.

### 5.3.3. *The tertiary sector*

English language teaching in the tertiary sector in France is a varied landscape. Language teaching for specialists takes place in Modern Language departments, where there are two types of undergraduate degrees: Foreign and Regional Languages (*Langues, littératures et civilisations étrangères et regionals*, or *LLCER*) and Applied Foreign Languages (*Langues étrangères appliquées*, or *LEA*). As there is no national curriculum for these degrees, the content varies enormously from one university to another. Most of the classes are taught either by associate professors with tenure (*maîtres de conférence*) or full professors (*professeurs des universités*) who may or may not have attended teacher training college, or by teachers who have either the CAPES or the aggregation, but were recruited to teach in the tertiary sector. A proportion of the classes are also taught by teaching assistants, or other contractual teachers from outside the university. Typically, *LLCER* courses offer phonetics classes in the form of lectures and seminars, especially in the first year, and sometimes in every year through to the second year of the Master’s courses, and most *LEA* courses, though not all, offer at least some phonetics, usually in the first year. The number of classes and the level of pronunciation required and achieved by students in some *LEA* departments can lead to feelings of linguistic insecurity, especially



when students measure their level against their peers on *LLCER* courses (Wharton & Wolstenholme, 2019).

Outside the language departments, all students in the tertiary sector are required to study a foreign language for the duration of their time at university, and language courses are also offered at doctoral schools. These courses of languages for specific and academic purposes (LSP / LAP) are referred to as *Langues pour spécialistes d'autres disciplines (Lansad)*, and mostly consist of English classes, although universities do offer other languages if students prefer. The quantity, quality, and content of English classes in the *Lansad* sector vary enormously (Taillefer, 2002), with relatively few research posts and a higher than average amount of courses taught by external teachers on short-term contracts (Braud et al., 2015: 21), and few classes lead to any form of certification.

Universities and other institutions in France increasingly offer courses partially or entirely taught in English, and the number of these English Medium Instruction (EMI) courses are growing yearly, partly because they appeal to international students, but also because they are seen as a top-down way of improving language skills amongst French students (Pagèze & Lagabaster, 2009). These classes are usually offered by L1 French teacher who usually get little or no help to teach in English; for the moment, only Bordeaux University offers an in-house dedicated training course for EMI teachers (Lasagabaster & Pagèze, 2018). As this area is rapidly growing, but research and research-based training in EMI is lacking in France, in 2020, I started an international research project with a number of colleagues from universities in France, Spain and the Czech Republic to investigate EMI in universities. The project, Interphonology - Czech, Arabic, French, English and Spanish (IP-CAFES) was an attempt to identify the phonological features responsible for comprehensibility and intelligibility problems which may contribute to cognitive charge in the students attending courses given in foreign-accented speech (Frost et al., 2021). The project entailed taking samples of continuous speech from lectures given in English by teachers with different L1 accents, and using PRAAT to tweak certain phonological features, making them either closer or further from a typical L1 English speaker's realisation. The samples were then to be played to a variety of listeners, who use *Idiodynamic* rating software to rate which parts are easier or harder to understand. A short self-confrontation interview then helps to better understand how certain features may prove a hindrance to comprehensibility. Unfortunately, largely due to the Covid-19 pandemic, this project has had to be paused for the time being, but the pilot tests I carried out last year were very promising.

Finally, some universities offer separate language courses for lifelong education, and some, such as Grenoble Alpes University, have a dedicated team of language teachers for this service, Grenoble Alpes University being one of the few, and it is in this service where I am currently employed. One of the

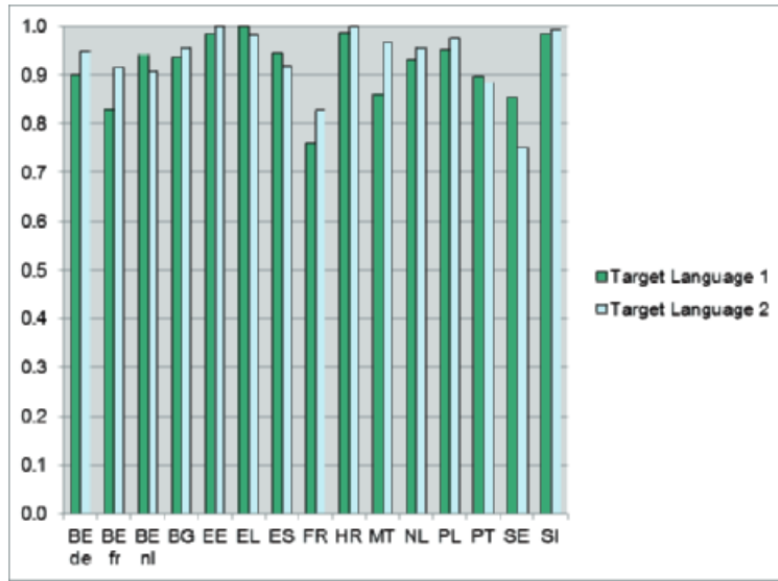
courses I teach is on academic oral English, to help colleagues to present their research in English at conferences. Some of the recordings made in the context of this course are being used in the learner corpus which I will present at the end of chapter seven.

#### 5.3.4. *The private sector*

In France, as in any other country, there are many private language courses on offer outside the state education sector, either in language centres, or by freelance teachers and trainers. I have employed many teachers from the private sector over the past twenty years, and most of them were not trained teachers. There are now several CELTA (Certificate in Teaching English to Speakers of Other Languages) courses in France: one run by the British Council in Paris, and there are also courses now in Strasbourg, Lyon and Brittany, however these are courses taught by private language schools; until twelve years ago, there was no alternative in the French state education sector to the CELTA. It was due to this lack of a French teacher training certificate aimed at English teachers in the private sector, that we decided to launch the *Diplôme d'Université de formation de formateurs et de formatrices en anglais* in 2010 at Grenoble Alpes University. I was part of the team that designed the course, I have run it since 2015, and we have now trained 150 teachers from France, the UK, America and many other countries. Although the course is primarily aimed at teachers in the private sector, some of the teachers have gone on to take the *CAPES* and now teach in secondary schools or universities. I teach several modules on the course, including one on teaching pronunciation, and the course also includes a module taught by a colleague on the use of the body and the voice, which draws heavily on the work of the THEMPPPO team, which I will present in chapter seven.

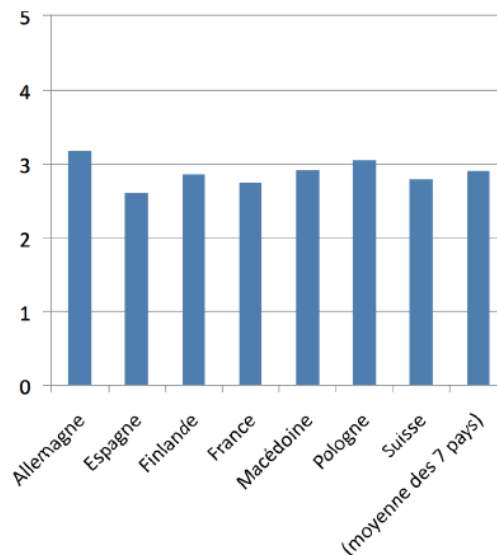
#### 5.4. Teacher training and pronunciation: facts and figures

The European Survey on Language Competences (ESLC) mentioned in section 5.2 does not ask teachers what and how they teach regarding pronunciation, however in an earlier item, they were asked about their teacher training, as we can see in figure 5.4a:



**Figure 5.4a. Received training in teaching the target language as a foreign language, ESLC (European Commission, 2012b: 196)**

The graph represented in figure 5.4a represents the proportion of teachers who have received training to teach the first and second target languages taught either in initial or in-service teacher training (with 1.0 equal to 100% of teachers): in France, this means English and Spanish. The graph shows that of the 12 countries concerned by this questionnaire, France is the country with the most poorly trained teachers in target language 1 (English). As we will see in the following sections, the picture in France is very mixed, but this finding does tally with some research that I carried out ten years ago, the English Pronunciation Teaching in Europe Survey (EPTiES) as part of a team of ten researchers representing language teaching in 11 different countries in the primary, secondary, tertiary and private sectors (Frost & Henderson, 2013). A total of 843 teachers responded to the questionnaire, composed of 84 items, and in France, the number was 65. We then followed up the questionnaire by conducting semi-structured interviews with a sample of the participants. The most interesting findings for me concerned the paucity of training that teachers had received to teach pronunciation. The following graph (figure 5.4b) represents the quality of teacher training which the participants reported that they had received to teach pronunciation.



**Figure 5.4b. Perceived quality of training to teach pronunciation. (Frost & Henderson, 2013: 102)**

The graph represented in figure 5.1b shows the responses to item 57: “In relation to pronunciation, please rate the teacher training you received from 1 to 5, with 1 as “extremely poor” and 5 as “excellent” (Frost & Henderson, 2013: 102). We can see that of the seven countries selected for the article from which this graph is taken (Frost & Henderson, 2013) that although the differences are small, teachers in France rated their training to teach pronunciation poorly, above only teachers in Spain. This figure is even more striking since 97% of the respondents in France were employed in public education, and so most of them had been through teacher training college. In fact, the paucity in training to teach pronunciation was in fact mentioned by teachers in all 14 countries in the survey (Henderson et al., 2015: 54). The follow-up question to item 57, which asked for detail on the nature of any professional training the respondents had received on teaching pronunciation. It became clear that many of the respondents who stated that they had received professional training to teach pronunciation had not, in fact: they had confused the professional training with undergraduate degree content, and they were in fact referring to the phonetics lectures and seminars they had attended during the first year of their undergraduate language degrees. In fact, of the 65 respondents in France, 19 reported having had little or no training in teaching pronunciation, 19 mentioned their undergraduate classes, and only 9 mentioned dedicated seminars or classes at teacher training college or elsewhere (Frost & Henderson, 2013: 103).

### 5.5. French as a foreign language (*Français langue étrangère*)

Institutionally, research and teaching in LX French, as for LX English, does not really have a natural home in French universities, in that there is no section in the *CNU (Conseil national universitaire)*, the

national body which regulates research posts in French universities, dedicated to this area of research. Research posts in this field, known in France as *la didactique des langues*, therefore come under the aegis of modern languages (the 14<sup>th</sup> section of the *CNU* for Romance languages, or the 11<sup>th</sup> section for English), so researchers also ask to belong to the Linguistics section (the 7<sup>th</sup> section of the *CNU*), a situation which complicates the careers of those working in the field (Savatovsky, 2011).

The teaching and research of pronunciation in LX French, be it in non-francophone countries (*Français langues étrangères*, or *FLE*) or to people who live in France or another country where French is an official language (*français langue seconde*, or *FLS*), is undergoing similar changes to those we have seen in the community of teaching and researching pronunciation of LX English. Since Daniel Coste and Robert Galisson co-edited their *Dictionnaire de didactique des langues* (1976) and Jean-Pierre Cuq edited his *Dictionnaire de didactique du français : langue étrangère et seconde* (2003), there have been many changes in this community, and a domain which was considered “new and original” had matured (Ferreira et al., 2010). Cuq himself stated that the field was now more autonomous, and leaned less heavily on fields such as linguistics (Cuq, 2005). And yet when it comes to teaching pronunciation, certain authors (Nocaudie et al., 2019) highlight the lack of formal knowledge in pronunciation and how to teach it amongst many LX French teachers. Clémentine Abel (2019) cites the infighting between proponents of the action-oriented approach of the CEFR, linguists and phoneticians as being largely responsible. French as a foreign language usually concerns adult learners and is a well-defined field, but less so French as a second language (Fenclová, 2014: 151), which is influenced by its perception as “the language of schooling” for (im)migrants to France (Cuq, 2003). This area is politically very sensitive, with linguistic policies influenced by the current French government’s line on integration rather than multiculturalism (Bauvois, 2019).

The idea that intelligibility may be a goal rather than a native-speaker model has become more prevalent in the past decade in France, perhaps more thanks to the Common European Framework of Reference for Languages (Council of Europe, 2001), whose descriptors mention intelligibility at B1 level (Didelot et al., 2019: 3), than to research published in international journals. However, certain notions which are largely rejected in International scientific journals are still part of the *doxa* of teaching pronunciation in LX French (Miras, 2019: 2), such as the importance of the native speaker model. Jérémie Sauvage (2019: 6) suggests that the epistemological tensions around teaching pronunciation are at least in part due to it being a neurophysiological process, unlike teaching grammar. Corinne Weber (2019) also calls for a new paradigm for teaching pronunciation in French, and joins Miras in emphasising the importance of recent research in socio-phonetics to inform the question of which model(s) to teach, such as the corpus project *Interphonologie du français contemporain* (Detey & Racine, 2012). Laura Abou Haidar also stresses the importance of oral corpora in defining learning

objectives for teaching oral LX French based on the many varieties of French spoken around the world (Abou Haidar, 2021).

Leaving aside for the moment the native-speakerism and intelligibility debate, let us turn to questions of how to teach pronunciation, and to the subject that is central to this report, prosody. Over the past two decades, there have been several researchers who underline the importance of prosody for teaching LX French, such as Philippe Martin (2019), who first pioneered the use of F0 curves for LX French learners in Canada in the 1970s (James, 1976), and later developed Winpitch, a digital tool for helping learners with French intonation (Martin, 2005). Grégory Miras also places a great deal of importance on prosody, and chose to use instrumental music as a pedagogical approach for teaching French intonation for his PhD study (Miras, 2014). Sophie Aubin also uses music in her work, and calls for a stable, durable and permanent approach to bringing together the theory and practice of musicology and that of teaching French, and emphasises the importance of working on the rhythm of the language in order to improve intonation (Aubin, 2022: 192). Aside from music, other performing arts have often been used in LX French, and without specifically making prosody or even pronunciation their aim, it is central to the approach used, for example the use of theatre (Payet, 2010), or Camille Vorger's use of slam poetry in class (Vorger, 2012).

However, for those teachers and researchers who are interested in pronunciation of LX French, and have found mainstream "corrective phonetics" to be lacking, they have looked elsewhere for teaching techniques and adapted them to teaching pronunciation. One example of this is the Silent Way (Gattegno, 1963), and this is used in France for teaching French, also English and other languages (Young and Messum, 2012), however it is an approach which focuses more on phonemes than prosody. Another notable example of an alternative pedagogical approach being used by LX French teachers is the Verbotonal System developed by Petar Guberina (Guberina, 2013). In the 1950s, Guberina developed this system for helping children and adults with hearing and speech pathologies, and his approach has been quite widely used in LX French teaching for several decades. The Verbotonal system seeks to improve pronunciation by re-educating the ear, and by mobilising the whole body, and it places a great importance on prosody (Borrel & Salsignac, 2002; Sauvage & Billières, 2019; Bourreux et al., 2021). Michel Billières (2016: 19) states that the Verbotonal system differs from what he calls "the articulatory method", which in his view ignores prosody, apart from the occasional listen-and-repeat exercises on intonation and rhythm. The Verbotonal system is not widely known about outside the community of LX French teaching and research, but a study is currently underway in Spain to use this approach to teach LX English to Spanish and Catalan speakers. The researcher, Stella Ville, is a French doctoral student who I am co-supervising. After discovering this method while teaching French

in Spain, she decided to apply it to teaching English (Ville, 2021), and will be joining the Lidilem laboratory at Grenoble Alpes University for a year as a visiting doctoral student in 2023.

## 5.6. Technology mediated pronunciation learning

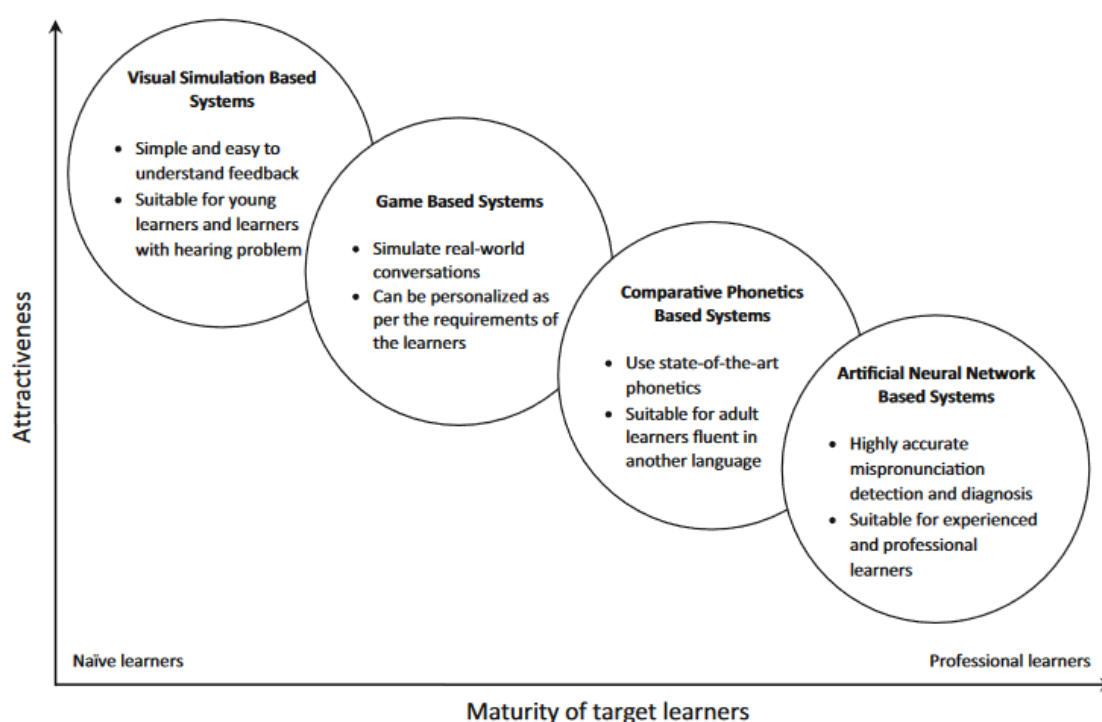
Language teachers have always been early uptakers when it comes to new technologies, especially those interested in teaching pronunciation. Alain Ginet, in the introduction to his edited volume (Ginet et al., 1997: 8), mentions that the first language laboratory in France, which used wax cylinders to record short extracts of speech in an early version of the audio active comparative approach, was piloted by Théodore Rosset, director of the *Institut de phonétique* in 1902, in Grenoble. Over the next hundred years, we have seen analogue technology gradually replaced by digital technology, and in the last two decades, the number of digital tools has greatly increased. The potential for what is generally known as Computer Aided Pronunciation Training (CAPT) is huge, and it is not the main focus of this report, but I would like to present some work, including some of my own research, which has focused on prosody.

I prefer the term “mediation” and I refer to “learning” rather than “training” in the title of this section because much of the pronunciation learning happens informally outside the classroom, and often outside the control of teachers and researchers. I also prefer the word “technology”, to “computer”, as many learners, especially teenagers and young adults, prefer to access English either for formal or informal learning using handheld devices, especially smartphones, via platforms such as Twitter and other social networking sites (Fouz-González, 2017). Geoffrey Sockett’s research on the Online Informal Learning of English (OILE) habits of French university non-English major students (Sockett, 2014) has shown gains in pronunciation due to listening to music (2014: 82-83), and to television series (2014: 130). OILE research has mainly explored lexical and grammatical gains from OILE, but Sockett and his colleagues make some suggestions for future pronunciation research (2014: 153-154). This is an area which I would like to explore more in my research, indeed Geoffrey Sockett and I have spoken about it on several occasions, but controlling for the many variables involved in OILE is challenging, especially regarding the prosodic features I would like to try to measure. In terms of teaching, however, I see OILE as a space where guided learning can happen, provided learners are trained in how to engage with the language they are exposed to. This means raising awareness to and practising the physical aspects of selected phonetic and phonological features, and developing various strategies to help learners to be more than simply passive consumers of entertainment in English.

When it comes to digital resources actually designed for pronunciation training, as Pennington and Rogerson-Revell (2019: 238) point out, “one of the difficulties is that there is no obvious fit between

language learning pedagogies and the affordances of digital technologies”. The theme of the 8<sup>th</sup> annual Pronunciation in Second Language Learning and Teaching (PSLLT) conference in 2016 was “The Role of Technology in L2 Pronunciation Research and Teaching”. In the introduction to the proceedings, Mary O’Brien and John Levis (2017: 1) warn however that many of the commercial CAPT tools available, attractive and promising though they may appear, are often “neither pedagogically sound nor informed by research” and call for more collaboration between researchers, software developers and classroom language teachers.

Based on a review of CAPT studies from 1972 to 2017, Chesta Agarwal and Pinaki Chakraborty (2019) class digital tools for pronunciation training into four types, as we can see in figure 5.7:



**Figure 5.7. Salient features of different categories of CAPT systems for English. (Agarwal & Chakraborty, 2019: 3740)**

In figure 5.7, the categories are arranged according to their attractiveness and the expected level of maturity of the target learners. Lynn Henrichsen (2019) suggests a set of guidelines for evaluating CAPT tools, and in a later paper (Henrichsen, 2021) she proposes the following taxonomy of CAPT tools, listing: text and audio only; listen and repeat; listening discrimination (minimal pairs); visual articulatory displays; visual acoustic displays; Automatic Speech Recognition (including speech-to-text recognition software); corpora. This is not to say that all digital tools fall into one or other of these categories: some digital tools use two or more of the systems presented in figure 5.7, incorporating visual tools, using gamification, listen-and-repeat type activities and artificial intelligence. An example



of this is the app Blue Canoe (Sakach, 2022), which uses a colour system for vowels, a variety of gamified activities, a listen-and-repeat function with Automatic Speech Recognition which recognises whether learners are using word stress correctly.

As this is such a vast field, I will limit myself in this section to a brief overview of some digital tools which have focused on prosody, and which I have either used myself or tested at conferences and workshops. There have been a number of attempts to represent prosodic features visually, often developed by researchers and inspired by research tools such as PRAAT. There have been many studies which show that various digital technologies can help improve prosody in LX learning, often using pitch contour visualisation (Anderson Hsieh, 1994; Hardison, 2004; 2013; Hamlaoui & Bengrait, 2013). In the context of English and French, there have been tools developed for visualising the intonation curve of French for L1 English users, such as *Winpitch*, mentioned in the previous section (Martin, 2005). Nadine Herry-Bénil and a team of developers from the *Laboratoire de la parole et du langage* in Aix-en-Provence developed a tool called *Prosodia*, for L1 French learners of English, which used an F0 visualisation interface along with a series of activities (Herry & Hirst, 2002). Rather than focus on intonation, Anthony Stenton chose to focus on the stress and rhythm of English with his L1 French learners. Stenton and his colleagues developed synchronised web authoring notation system (*SWANS*) which provided a visual aid for oral comprehension in the form of synchronised subtitles annotated for prosody and used audio and video files (Stenton & Gimeno, 2011).

My first direct experience of using digital tools to research pronunciation learning was using Jean-Claude Bertin's *Learning Labs* programme (Bertin, 2000). In a small action-research study I carried out for my Master's study (Frost 2002), I used *Learning Labs* to enable high school students to visualise the intonation and stress of their own speech output when they recorded themselves repeating words and short phrases chosen from a text they had been studying. The idea was that one of the algorithms in *Learning Labs* would help them to better visualise the stress placement in their recorded utterances, and compare it to the original recording – they were to repeat the process until the two curves were identical. Needless to say, they never achieved this goal, which actually led to a positive result, as they made many recordings. It was, I believe the high number of repetitions rather than the F0 visualisation tool which led to the small gains in improvement. This was also the conclusion of Cazade (1999), who suggested that even the wave forms and F0 visualisation tools available to learners in the nineties could prove useful, provided that the learners had some explicit phonetic training, and were able to engage actively with the program they used. This study therefore convinced me that technology could help learners to improve their pronunciation and comprehension, but that digital technologies were an aid, and not a substitute for classroom teaching.

For my PhD study, I continued to investigate the use of technology to improve the pronunciation and oral comprehension skills of French learners, by building an interactive web-based tool to work on word stress using authentic audio recordings of a variety of L1 and L2 English speakers (Frost, 2008). While it is true that digital technology has greatly improved in the last twenty years, digital tools are still constrained by the technology of the day, in a way that face-to-face teaching is not. Perhaps for this reason, as Rogerson-Revell (2019) states:

many CAPT resources are less innovative pedagogically than one would expect. Indeed, in some cases, as technology progresses, pedagogy appears to regress, returning to audiolingual approaches of repetition, mimicry, and drilling. While such methods still have their place, they are not sufficient to help develop communicative or phonological competence in a language. (Rogerson-Revell, 2019: 191)

I still believe that pronunciation, and in particular prosody, must be *felt* for learners to truly progress, both in terms of reception, and in terms of production. Digital tools used in the classroom or informally outside the classroom may help to improve learners' pronunciation, and are often a useful complement to classroom teaching. As Jonás Fouz-González (2015) reminds us, guided self-monitoring is important if learners are to maximise their potential, and the role of the teacher in this space is more of a guide, or a mediator between the learner and the digital environment, which is a role that a teacher often plays in any communicative approach to language teaching (Chini, 2001: 3).

In 2021, the Covid-19 pandemic forced many teachers to move their classes partially or wholly online, either synchronously or asynchronously, with varying degrees of satisfaction and success. An example of a successful shift of a pronunciation course online was published recently by Leticia Quesada Vázquez (2021; 2022). The two pronunciation courses in question, for English undergraduates in a Spanish university, were positively evaluated following the sudden and unexpected switch to Synchronous Computer-Mediated Communication (SCMC). Leticia Quesada Vázquez, completed her doctoral thesis focussing on teaching the rhythm of English to Spanish in 2019, in the Universitat Rovira i Virgili (URV), Tarragona, Spain. I sponsored her as a visiting doctoral student in the Lidilem research laboratory at Grenoble Alpes University in 2018-19, and she successfully completed her PhD and continues her research at Universidad Nebrija, Spain, where she is director of the Modern Languages bachelor's programme.

In this chapter, we have seen that there are many environmental factors which contribute to the process of LX learning. In accordance with Bronfenbrenner's ecological systems model (Bronfenbrenner, 1974; 1977; 1994), and within an embodied cognitive framework (Clark, 1999; Varela et al., 1992; Wilson & Golonka, 2013), teaching and research should take account of external environmental factors. Although learners, teachers and researchers may not have direct control of many of these environmental factors, they often provide opportunities and constraints which may have influenced the learning process, and therefore we must be aware of the wider environment in

order to improve learning. By understanding the wider environment, teachers can help learners to monitor their own learning and to develop appropriate strategies. Political decisions, technological developments, industrial action, pandemics: for researchers, there are many variables which are impossible to control, but possible to control *for* when attempting to measure learning in the field of LX pronunciation. In the next chapter, we will examine the three most frequent constructs in this field, namely accent, intelligibility and comprehensibility.

## CHAPTER SIX

### Prosody in production, perception and comprehension: LX research

As we saw in chapters one and two, we know from a large body of research dating back to the 1950s that prosody is essential for perception, segmentation and comprehension of speech, as summarized by Anne Cutler and Delphine Dahan (Cutler & Dahan, 1997), in fact, this is the case not only for the prosodic features of a word under analysis, but also features several syllables away (Dilley et al, 2010). This chapter is concerned with prosody on the context of cross-linguistic studies, and therefore with linguistic transfer, both at the level of perception, and at the level of production.

At the level of perception, we saw in section 3.4.2 how listeners' L1 prosody can act as a filter, and cause problems perceiving and understanding speech in another language. It has also been shown by numerous studies that learners are able to recognise their own language and distinguish it from others by using prosodic cues only, that is by listening to modified speech samples, stripped of their phonemic information by using low-pass filters (Ohala & Gilbert, 1981). Incorrect prosody makes processing speech harder, as a number of recall experiments in the 1960s and 1970s showed (Martin, 1968; 1979; Leonard, 1974). A number of studies have also found longer reaction times for prosodically incorrect or misleading utterances (Meltzer et al, 1976; Phillips et al., 2022). It has also been demonstrated that the way we group language into metrical units when listening is influenced by the rhythmic structure of our L1 (Bolton, 1894; Woodrow, 1951; Fraise, 1963, 1982; Iverson et al., 2006; Arviniti, 2020).

At the level of production, LX prosody transfer can be measured using acoustic analysis of the measurable physical correlates of prosodic features, but it is most frequently measured using listener tests (sometimes, rather confusingly, referred to as perception tests). Listener tests are the most common measure in LX pronunciation instruction studies, and the chief constructs which are used in these listener tests are accent, intelligibility and comprehensibility.

I will therefore begin this chapter by exploring how these constructs are defined. Accent is a well-defined construct, but intelligibility and comprehensibility - as they are widely understood by the LX pronunciation teaching research community - are not, and this is problematical both for research and for teaching. I will therefore present definitions of intelligibility and comprehensibility drawn from the clinical community, as I believe this definition is clearer, and makes assessment of these constructs easier, which in turn can improve instructional design, as we shall see in the next chapter. I will then explore how accentedness, comprehensibility and intelligibility have been studied in the LX

pronunciation community, with the aim of exploring how prosody and segmental features contribute to these constructs. Then we will examine studies from the past fifty years on learning outcomes of pronunciation instruction which focuses on prosodic features. This literature review will concentrate on prosodic features related to stress and intonation, but also on other suprasegmental features such as pauses, which help to break up continuous speech into prosodic units such as tone units. Studies range in scope from targeting a single cue of one prosodic feature, to targeting multiple segmental and suprasegmental features in the same study. Studies may be more production-oriented, or more perceptual in nature, or they may rely on both production and perception. In the final section of this chapter, I will present a model of language use in an embodied cognitive framework, and how intelligibility and comprehensibility figure in the process of oral communication.

### 6.1. Accent, intelligibility and comprehensibility

In their article entitled “The foundations of accent and intelligibility in pronunciation research”, Murray Munro and Tracey Derwing (2011) review research from applied linguistics and L2 pronunciation teaching from the seventeenth to the twenty-first century. Munro & Derwing’s influence in the field of L2 pronunciation teaching research cannot be understated, and the way most researchers understand and define the constructs of accent, intelligibility and comprehensibility aligns with their work, notably with their article published nearly thirty years ago: “Foreign accent, comprehensibility, and intelligibility in the speech of second language learners” (Munro & Derwing, 1995).

Okim Kang and her colleagues (2021: 32) point out that both segmental and suprasegmental features have been linked in studies which focus on the three constructs of accent, intelligibility & comprehensibility, and as we saw in section 1.1.6, there is unequivocal evidence of the importance of suprasegmental features for lower order processes, such as segmentation, as well as for higher order cognitive processes such as understanding pragmatic features. In a study conducted over thirty years ago, Janet Anderson-Hsieh and her colleagues found prosodic variables to be the most strongly associated with judgements of accurate pronunciation regardless of the language subgroup under investigation (1992: 548), as they pointed out, however:

More research is needed that investigates the relationship between native speaker judgments of pronunciation and deviance in segmentals and prosody. Whereas more evidence has been presented to support the primacy of prosody over segmental in impressionistic ratings of nonnative pronunciation, the results of the studies supporting this were not strongly conclusive. (Anderson-Hsieh et al., 1992: 534)

There has indeed been much research investigating the roles and relative importance of various phonological features over the last thirty years. We will examine a cross-section of that research over this chapter, beginning with some definitions of the three most frequently investigated constructs.

### 6.1.1. Defining the constructs

In pronunciation research, the construct of **foreign accentedness** seems so intuitive, that neither Derwing and Munro (1995) nor Derwing and Munro (1997) offer a definition. As we saw in section 3.2.1, David Crystal's definition of **accent** places importance on both the speaker *and* the listener: "The cumulative auditory effect of those features of pronunciation which identify where a person is from, regionally or socially" (Crystal, 2008: 3). The *Chambers Dictionary* defines **accent** as "any way of pronouncing speech characteristic of a region, a class or an individual" (Chambers, 1993: 8), which also emphasises the role of the listener to a certain degree. Whereas the more "phonetically-oriented literature" may define foreign **accentedness** in terms of the divergence of measurable acoustic properties of speech output from a given norm (Thomson, 2017: 3), Munro and Derwing necessarily define accent in terms of both production and perception, as their research uses perception studies in various forms. We will define (foreign) accentedness here as follows:

**Foreign accent** may be defined as a set of acoustic and perceptual phonetic characteristics present in the speech of an LX language user which are recognisably different from those used by L1 speakers of that language, and which are common to language users belonging to another language community.

As with the construct of accent, definitions of **intelligibility** also place the onus on both speaker and listener. For most people outside the domain of pronunciation research, the terms **intelligible** and **comprehensible** are synonymous. The *Chambers Dictionary* defines intelligible as "capable of being understood, comprehensible, easy to understand" (Chambers, 1993: 870). Pre-1995, these two terms were often interchangeable, even among researchers in pronunciation. For John Catford (1950: 8), an utterance is **intelligible** "if the hearer understands the words, i.e. if his response is appropriate to the linguistic forms of the utterance". Susan Gass & Evangeline Veronis (1984) carried out an influential study on what they termed **comprehensibility**, where this construct was measured through successful word or utterance recognition. However, since Munro and Derwing's 1995 study, a consensus has been reached in the field of pronunciation research as to what intelligibility and comprehensibility mean, despite the terms not being clearly defined as such, but rather defined by default through the measurement tools used in early studies. **Intelligibility** is generally understood to mean "the extent to which a speaker's message is actually understood" (Munro and Derwing, 1995: 76). In order to measure **intelligibility**, some task or tasks must be used which show to what degree listeners understand the intended meaning of the speaker, for example orthographic transcription tasks, or questions to check comprehension. There is also a consensus among researchers as to the definition

of **comprehensibility**, which again follows the measurement techniques used by Munro and Derwing (1995). Munro and Derwing's 1995 article provided no definition of **comprehensibility**, but a *de facto* definition was arrived at by their measurement technique, which used a Likert scale to express *perceived ease of* understanding: 1= extremely easy to understand and 9 = impossible to understand. They explicitly defined **comprehensibility** in a later article (Munro & Derwing, 1997: 2) as “judgments on a rating scale of how difficult or easy an utterance is to understand”.

There is no doubt that according to the above definitions, comprehensibility is a much easier construct to measure than intelligibility, but a figure between one and nine based on listener perception of ease of comprehension is of limited use to language professionals. However, measuring intelligibility involves tasks such as orthographic transcription and comprehension-checking questions, which amounts to measuring what the CEFR (Council of Europe, 2001) refers to as “oral comprehension”, or “listening”, or what many consider “the most difficult skill to learn” (Martínez-Flor & Usó-Juan, 2006). Techniques for measuring both intelligibility and comprehensibility demand the active participation of a listener, but tasks conceived to measure the construct of intelligibility in particular imply a complex set of rater skills, and intelligibility measured in this way also reflects on more than just the speaker's pronunciation, since it may also be influenced by lexical and grammatical knowledge, and other factors, such as speech rate (Derwing, 1990; Munro & Derwing, 1998). The difference between “listener” and “rater” is not always clear, but in pronunciation teaching research, the term “listener” is usually used to refer to people who listen to speech samples and make judgements on *comprehensibility* (Kennedy & Trofimovich, 2019: 275), and “rater” is used for people who use a more complex set of skills to make judgements on other factors, such as fluency, intelligibility, level, etc. (Isaacs & Trofimovich, 2022). As Kang & Ginther (2018: 3) state: “listeners become raters when they are trained to rate to a scale”.

### *6.1.2. Another approach to intelligibility and comprehensibility: clinical communities*

The definition of intelligibility and the methods of measuring it are not the same once we leave the field of pronunciation research in applied linguistics and teaching. In medicine, the branch most concerned by pronunciation is speech therapy, and I believe it may be helpful to our understanding of this construct to see what intelligibility means to speech therapists. For his PhD, Nicholas Pommée (2021) explored different definitions of intelligibility and comprehensibility within his field – medicine - in order to better understand clinical solutions to problems of intelligibility. The approach he adopted was a “Delphi consensus study”, which was developed in the 1950s and 1960s by Norman Dalkey and Olaf Helmer in an attempt to gain reliable expert consensus. They named their approach after the

Ancient Greek *Oracle of Delphi* because they wanted to promote anonymity and avoid direct confrontation between experts, and thereby achieve a gradual consensus of opinion on topics which could aid independent thought and decision-making (Barrett & Heale, 2020). Prior to launching the survey, Pommée carried out a review of definitions, which revealed that a certain level of consensus already existed in the clinical field as to definitions of the two constructs (Pommée et al, 2022: 5). He used this as a basis for questioning a panel of 40 experts from different fields, including clinicians, researchers, lecturers, and other industry professionals who were engaged in activities in speech sound disorders and/or fluency disorders (Pommée et al, 2022: 4). The process comprised three rounds: the first round contained 31 questions, the second questionnaire used Likert scale items to react to 22 statements and the third round comprised three statements. The questionnaires included items on definitions and synonyms of the two terms, production and perception of language, and methods for measuring speech. The results of the study are very interesting, both in terms of the definition of the two constructs of intelligibility, and how they should be measured. The results of the study for intelligibility are as follows:

**Intelligibility** refers to the reconstruction of an utterance at the acoustic–phonetic level, intelligibility-related information is thus carried by the acoustic signal (i.e., intelligibility focuses on signal-dependent information). This reconstruction is made possible both by the speaker’s phonetic–acoustic production ability and by the listeners acoustic–phonetic decoding skills. (Pommée et al., 2022: 11)

As to the assessment of intelligibility, recommended methods involve the participation of a listener, but according to very selected and guided tasks, and importantly, the importance of linguistic factors other than phonetic and phonological factors should be limited:

Perceptually, intelligibility is best analysed on low-predictability stimuli: phonemes, syllables, pseudo-words, but also words (in minimal pairs) and unpredictable sentences for a more functional assessment taking coarticulation and phrase-level symptoms into account (e.g., respiration and prosody), as long as top-down cognitive compensation processes of the listener are avoided (i.e., no help from semantic or linguistic context) (Pommée et al., 2022: 11)

Comprehensibility is, according to the results of this survey, a more complex process, involving higher-level cognitive operations:

**Comprehensibility** refers to the reconstruction of a message at the semantic–discursive level, subsequent to the acoustic–phonetic reconstruction. Therefore, intelligibility is a component of comprehensibility. In addition to the acoustic–phonetic decoding, it also includes signal independent, contextual elements such as the linguistic or the non-verbal context. (Pommée et al., 2022: 11)

When measuring comprehensibility, other linguistic factors may be taken into account:

Comprehensibility refers to the more functional dimension of communication and is perceptually best assessed using meaning-related ratings (i.e., taking into account top-down cognitive processes that might compensate for degraded acoustic–phonetic information). (Pommée et al., 2022: 11)



These definitions concur to a large degree with definitions of intelligibility and comprehensibility to be found in the literature in the field of clinical speech disorders, for example Kathryn Yorkston and her colleagues' definitions (1996), taken from an article on dysarthric speech:

The term intelligibility refers to the degree to which the acoustic signal (the utterance produced by the dysarthric speaker) is understood by a listener. [...] The concepts of comprehensibility and intelligibility may be distinguished by the fact that comprehensibility incorporates signal-independent information such as syntax, semantics, and physical context. (Yorkston et al., 1996: 55)

The definition of comprehensibility which is shared by the clinical community also tallies with the definition of Saito, Trofimovich, and Isaacs (2015), which goes further to exploring the linguistic components of comprehensibility than previous studies on the subject, and which I will present in the section 6.1.3. Clinicians working with speech disorders such as dysarthric speech, stroke-related aphasias, and other pathologies do not have the same concerns as language teachers and researchers, and of course improving intelligibility and comprehensibility may involve very different procedures in the clinical profession and in teaching. However, I believe that the above definitions can benefit LX pronunciation researchers: in this paradigm, intelligibility is more pronunciation-oriented, whereas the approach to measuring intelligibility in LX pronunciation research is actually closer to what clinicians would call comprehensibility. If we can find a way to measure intelligibility in LX research which brings us closer to pronunciation, this would help to develop resources, techniques, and learning objectives for teaching pronunciation.

### *6.1.3. Accentedness in LX research*

As we saw in section 6.1.1., foreign accentedness is a combination of segmental and suprasegmental features, however different researchers have focused on different features of perceived accent, in order to better understand how they contribute to our understanding of accentedness. In Munro and Derwing's classic 1995 study on foreign accentedness, a group of L1 English listeners were asked to rate accentedness in a series of recordings of 10 proficient LX English speakers whose L1 was Mandarin, (1 = no foreign accent and 9 = very strong foreign accent). They were then able to see whether these findings correlated or not with perceived comprehensibility (a similar 9-point Likert scale) and intelligibility, measured using an orthographic transcription task. They found no clear relationship between intelligibility and comprehensibility (1995: 94), and "although strength of foreign accent is correlated with perceived comprehensibility and intelligibility, a strong foreign accent does not necessarily reduce the comprehensibility or intelligibility of L2 speech" (1995: 74). This study was influential partly because it created a repeatable set of testing procedures, and also because of its central finding that there did not seem to be a correspondence between perceived foreign accent and

either intelligibility or comprehensibility. Munro and Derwing (1995: 94) suggest more research with different accents, different levels of proficiency, and of rater differences would be necessary to “elucidate the relative contributions to intelligibility of specific elements (subsegmental, segmental, prosodic) of pronunciation.”

Several studies have examined the “contributions of specific elements” to accent - although there is more interest in factors contributing to intelligibility and comprehensibility. By using digital processes to modify some or all of the acoustic parameters of speech, typically of the nucleus of stressed or unstressed syllables (F0, duration, amplitude, and formant structure), researchers can see how these features affect the perception of accentedness and comprehensibility and intelligibility. Harriet Magen (1998) investigated the role of segmental features, lexical stress and phrasal stress in the perception of Spanish-accented English by L1 English speakers, using modified speech samples and phonetically-trained raters, so she could explore certain phonemic features typical of Spanish accents (epenthetic schwas, vowel reduction, and fricative voicing). The results of her study show that suprasegmental features (stress), contribute consistently to judgements of accent. The segmental factors which contributed to accent perception - specifically “vowel reduction” and “tense-laxness” were also considered as suprasegmental elements - because duration was manipulated and these factors contribute to timing.

Duration has often been studied, partly because it contributes to the rhythmical structure of language, but also because it is one of the easier parameters to manipulate. Philippe Boula de Mareüil & Bianca Vieru-Dimulescu (2006) investigated the effects of duration, or timing, compared it with the correlate of F0 in a set of two experiments using prosody transplantation between Spanish and Italian. Their results showed that listeners were more influenced by prosody in their decisions regarding accentedness. The second experiment showed that the listeners were equally influenced by prosodic and segmental cues (Boula de Mareüil & Vieru-Dimulescu, 2006: 264). Lieke von Maastricht conducted a series of experiments on Dutch learners of Spanish and vice versa for her PhD, using spontaneous and manipulated speech. She found prosodic transfer took place in both sets of speakers in production, however at the level of perception, pitch accent did not prove to be a useful cue. This is probably because of her choice of using semi-spontaneous speech, where pitch accent is less salient than in read-aloud stimuli (Maastricht, 2018: 180). She also found that perception of foreign accent was affected by the individual manipulations of intonation and speech rate used (Maastricht, 2018: 181). Elisa Pellegrino’s (2012) study using read-aloud texts showed that both segmental and suprasgemental features played a role in accentness detection (Pellegrino, 2012: 266). In a more recent study by Pellegrino with Sandra Schwab and Volker Dellwo (2021), duration was again the most important factor in accent detection, this time with L1 Italian users’ perceptions of German-accented Italian and

L1 Italian samples (Pellegrino et al, 2021). Okim Kang's (2010) study used a similar approach to Pellegrino, measuring acoustic features of speech samples to explore the relative salience of various segmental and suprasegmental features on judgements of accentedness and comprehensibility of L1 Chinese teaching assistants' English by L1 American English listeners. Kang found accent ratings were best predicted by overall pitch range, followed by proportion of stressed words to the total number of words, pause duration, and articulation rate (Kang, 2010: 310), whereas comprehensibility scores were mostly associated with speaking rates, however these measures were less significant, and pauses had no discernible effect. (Kang, 2010: 311-312).

Bianca Vieru and her colleagues conducted a series of experiments comparing the perception of foreign accent in French by L1 French speakers to automatic speech recognition of various features that contribute to accent (Vieru et al., 2011). The samples under investigation were French from speakers of Arabic, English, German, Italian, Portuguese and Spanish, and they also used L1 French speaker samples. Unsurprisingly, the results were different according to the accent in question. The perception tests showed that for certain languages and certain listeners, segmental cues were more important for accent identification, for example [v] instead of /b/ and [s] instead of /z/ for Spanish speakers, [i] instead of /e/ for Arabic speakers, but listeners also report using prosodic cues for speakers of different origins (Vieru et al., 2011). Marie-José Kolly and her colleagues also manipulated temporal correlates and formant structure of a set of samples of L1 German speech, modified either by transplanting French-accented German or English-accented German (Kolly et al., 2017). They found that listeners were able to identify both accents by temporal cues, and that accent identification was improved by cues in the formant structures. They also found that the presence of one phoneme in particular, the uvular /r/, made identification easier (Kolly et al., 2017: 131).

From the studies reviewed in this section, it is evident that both segmental and suprasegmental features contribute to accent perception; however, when studies isolate different features, the picture becomes more complicated. Of all the acoustic correlates to perceptual cues, duration and F0 seem to be the most important. The overall impression is one of interdependence: prosodic features are always present to some extent in predicting accent, and when segmental features contribute, they often do so in combination with prosodic features, notably linked to the rhythm or timing of the language or languages under investigation: when segmental features contribute to intelligibility, the correlate is formant structure, and this is strongly linked to syllable reduction, a factor linked to the prosodic feature of stress. Speech rate is also a factor in accentedness, though it has little or no impact on intelligibility. Finally, the effect of context is not to be overlooked. Segmental difficulties are specific to the languages concerned by the studies.

#### 6.1.4. *Comprehensibility in LX research*

I have already discussed Munro and Derwing's (1995) study using Mandarin-accented English which found that "a strong foreign accent does not necessarily cause L2 speech to be low in comprehensibility or intelligibility" (Munro & Derwing, 1995: 91). It may be argued that accentedness and comprehensibility are constructs which are close in nature, one being a perception of foreign accent, and the other being a perception of the effect of a foreign accent. Furthermore, both are usually measured with a nine-point Likert scale, and the result is therefore limited: for this reason, the most interesting studies are the ones which seek to explore a variety of linguistic components of comprehensibility.

In a study which predates the other studies in this section and which I mentioned in chapter three, Fayer and Krasinsky (1987) used samples of Puertorican Spanish-accented English which were rated by L1 Spanish users and L1 English users for "intelligibility, grammar, pronunciation, intonation, wrong words, voice, hesitations, distraction and annoyance". I will return to this study in section 6.2, but it can be said to concern comprehensibility, because it measures what Fayer and Krasinsky were referring to as "intelligibility" with a five-point Likert scale, and they considered many linguistic factors. They state that "intelligibility is hearer-based" and that "the total communicative effect of a nonnative message" has non-linguistic sources (such as "style, speed and hesitation") and linguistic sources ("how close or deviant the form of the message is to or from the target language") (Fayer & Krasinski, 1987: 313). They use the term "comprehensibility" as a synonym for intelligibility, but differentiate between objective and subjective measures of what they appear to see as the same construct: "Subjective designs involve evaluative judgments and some type of rating scale, while objective tests use operative interpretation of the deviant message, such as rewriting or rephrasing" (Fayer & Krasinski, 1987: 314). This notion of objective or subjective assessment of these constructs is something we see in the clinical community – Pommée refers to "perceptual and objective assessment methods of intelligibility and comprehensibility" (Pommée et al., 2022: 26). The features which most contributed to what Fayer and Krasinski call "intelligibility" were pronunciation and hesitation, and they conclude by suggesting that "if intelligible speech is the goal, this finding, in addition to the others discussed above, has several implications. ESL curricula should be reviewed to determine if pronunciation is given sufficient emphasis" (Fayer & Krasinski, 1987: 322).

Pavel Trofimovich and Talia Isaacs carried out a set of experiments which are of particular interest here, because they used a corpus of samples taken from 40 French speakers of English with different proficiency levels, albeit from Quebec, not France. (Isaacs & Trofimovich 2012; Trofimovich & Isaacs,

2012). These studies are primarily interesting because they seek to find the relative contribution to accent and comprehensibility of different linguistic variables in the domains of phonology, lexis, and discourse structure. The precursor study (2012) investigated 19 speech measures in three categories: phonology, fluency, and linguistic resources (Isaacs & Trofimovich, 2012: 483-485). Results from this initial study showed that although both accent and comprehensibility are linked to many speech measures, accent was “uniquely linked” to phonological features, including rhythm, and segmental and syllable structure accuracy, but comprehensibility was mainly linked to grammatical accuracy and lexical richness. (Trofimovich & Isaacs 2012: 905). The later study (Saito et al., 2015), used the same corpus, and clarified a picture which was already emerging: that accentedness judgements were mainly a result of vowel/consonant errors and word stress (Saito et al., 2015: 1), whereas comprehensibility is a much more complex construct.

The research reviewed in this section, especially the more recent studies carried out by Trofimovich, Isaacs and Saito, brings the pronunciation research and language teaching community closer to the definition of comprehensibility shared by clinicians working in the field of speech which we saw in section 6.1.2, i.e. “comprehensibility incorporates signal-independent information such as syntax, semantics, and physical context” (Yorkston et al., 1996: 55).

#### *6.1.5. Intelligibility in LX research*

As we saw in section 6.1.1, intelligibility is defined by the community of researchers in LX pronunciation as “the extent to which a speaker’s message is actually understood” (Munro and Derwing, 1995: 76), which means studies use some measure or measures of comprehension, such as orthographic dictations, or alternative forced choice protocols. I will begin by presenting research which focuses on different aspects of prosody, before examining some of the research into segmental features, and the interplay between both levels of cues.

An oft-cited study, and the first to clearly establish a link between missing or misplaced primary stress and intelligibility, was the work carried out by Laura Hahn (2004), previously mentioned in section 3.3. Hahn used speech samples taken from three recordings of an academic lecture made by an L1 Korean teaching assistant, one version with correct stress marking, one where primary stress was misplaced and one where primary stress was absent (Hahn, 2004: 206). By primary stress, Hahn does not mean primary lexical stress, but focus - as she explains: “in English, new and contrastive information is presented in stressed elements, and old or given information is expressed in unstressed elements. For convenience, these relationships can be called the given-new stress connection (GNSC)” (Hahn, 2004: 202). Hahn’s results showed that correctly placed primary stress enhances comprehension (Hahn,

2004: 215). Beth Zielinsky (2006; 2008) has also highlighted the importance of stress to intelligibility. Her first study was a case study of the speech of one Vietnamese user of LX English and she used three raters, and her follow-up study used a Vietnamese, a Korean and a Mandarin speaker, and three naïve Australian L1 users for an orthographic transcription task. In both studies, all three listeners relied heavily on the syllable stress patterns and segments in the speech signal to identify the speaker's intended words, and non-standard segments seeming to play a greater role in reducing intelligibility than did non-standard syllable stress patterns (Zielinski, 2006: 34-36). Zielinski also notes the complex interplay between segments and stress patterns: when mispronounced segments misled listeners, it was most often when they occurred in strong syllables (Zielinski, 2008: 71).

Another approach to investigating the role of stress in intelligibility is by examining the unstressed – rather than stressed syllables – after all, it is the perception of so many unstressed syllables which is often problematic for learners of English, as we saw in section 4.1.2. Bettina Braun and her colleagues found that Dutch-accented English is not harder for L1 English speakers to understand, but there are sometimes comprehension problems where the language-specific implementation of lexical stress differs between Dutch and English, notably involving schwa (Braun et al, 2011: 376).

Stress and rhythm have also been investigated using perception tests involving a forced choice protocol. Keiichi Tajima, Robert Port and Jonathan Dalby (1997), who investigated the role of duration in Chinese-accented English and L1 American English as perceived by L1 American English listeners. Their results showed that listeners were significantly more likely to recognise the words spoken with “native-like temporal properties” (Tajima et al., 1997).

As we have already seen, pausing can also be considered a prosodic feature (Kang, 2010). A pause may either be filled or empty (Crystal, 2008: 355), and pauses are usually studied as a measure of fluency, rather than of intelligibility (Wood, 2016). In a dual task protocol similar to that of Hahn's primary stress study, (2004), Sadi Phillips and her colleagues (2022) investigated whether non-target like thought grouping patterns in LX speech increase processing difficulty for L1 listeners, and if so, how. 34 L1 English speakers listened to three versions of 30 sentences spoken by a proficient L1 French speaker of English: one “authentic” version, with inserted pauses at the clause boundary between tone groups, and one with “run-on” sentences. Listeners had to respond true or false for each sentence, and reaction times to randomly inserted tones were also measured (Phillips et al., 2022: 3-4). As they predicted, non-target-like pausing resulted in slower tone detection reaction times compared to target-like pausing (Phillips et al., 2022: 8).

Segmental features can of course also have an influence on how well speech is understood, however the picture is complex, with research often revealing the interplay between segmental and

suprasegmental features, or what Zielinski (2006) calls “the intelligibility cocktail”. Takeki Kamiyama and Shigeko Shinohara (2010) investigated the influence of intonation patterns on the correct identification of word pairs in French by Japanese learners. The word pairs chosen were differentiated by the presence or absence of a vowel, creating oppositions between CCV and CVCV sequences. The target words were embedded in declarative or interrogative sentences read by an L1 French speaker and played to Japanese learners in a series of three experiments. The results showed that learners tended to respond better to target sequences with a higher pitch (Kamiyama & Shinohara, 2010: 2), showing that segmental distinctiveness can be enhanced by intonation. Durational prominence had a biasing effect, which the authors hypothesis is due to the final lengthening phenomenon of French (Kamiyama and Shinohara, 2010, 5). Okim Kang and her colleagues’ (2018) study also demonstrates the complex interplay between segmental and suprasegmental cues. Kang used Chinese-accented speech in her study, and the listeners were representative of each of the three circles of Kachru’s model of World Englishes. The results showed that depending on the measure used, different cues had varying degrees of importance. In the filtered sentence test, the most important cues were vowels and consonants. However, results for the true/false test showed that segmental errors did not significantly predict the speaker’s intelligibility, and listeners relied more on stress, intonation, pauses and speech rates (Kang et al., 2018: 137). Finally, in a recently published study, Page Wheeler and Kazuo Saito (2022) investigated the contribution of iconic gesture, visual speech and phonological accuracy to intelligibility. The results showed that the vowel errors reduced intelligibility and that iconic gesture increased intelligibility for L1 listeners.

The studies presented in this section point towards the importance of stress for intelligibility, and in particular durational cues. Pauses may also be considered as features of the metrical or rhythmical structure of language, and they also play an important part in intelligibility of spoken language. However, spoken language is not just a matter of auditory cues to intelligibility. An aspect of intelligibility which is often overlooked in research is the visible cues to language (facial expression, gestures, and visible movement of articulators). As we saw in section 1.1, visible cues are important in identifying speech components (Guellaï et al, 2014).

## 6.2. Teaching prosody and learning outcomes

There have been many studies showing the efficacy of targeted pronunciation training – even over short courses – and in this section, I will concentrate on studies which focus on prosody. I have analysed 32 studies (table 6.3) and sets of studies conducted over the last five decades, which explore the

effectiveness of focusing on prosody in language teaching, and I will briefly present those studies and their findings in the following three sections.

In recent years, there have been two meta-studies published on research in teaching suprasegmentals features, both of which ask specific questions of the literature in the field. The first, written by Mark McAndrews (2019), is particularly concerned with the effect of short periods of instruction, and the instruction time on the 17 studies he examined varies from one to six hours. McAndrews whittled down his original pool of studies from 50 studies to the 17 which he retains in his metastudy according to criteria based on sample size, whether or not they included instruction, whether or not they contained tests of both listening and speaking skills, and whether they included enough information for his calculations as to effect (McAndrews; 2019: 154). He also averaged effect sizes by the type of measurement used in each study (McAndrews, 2019: 154). His analysis also showed that suprasegmental instruction can benefit learners who have already had “substantial exposure to the target language”, and that even short periods of instruction can have a positive effect (McAndrews; 2019: 156). McAndrews’ strict criteria necessary for the calculation process he used meant that the number of studies he chose was limited to studies ranging from 1999 to 2016, eight of them featured Mandarin as the target language, three featuring Japanese, and only two studies featuring English.

Xue Wang’s (2022) metastudy, which compared the effects of teaching suprasegmental features and segmental features on intelligibility was more modest in scope, as she only considered English as a target language. The total number of studies included by Wang is therefore 14, ranging from 1997 to 2017. Instruction periods ranged from four to 17 hours, over one to 13 weeks, and the studies vary in focus, some including only segmentals, some comparing the effects of teaching segmentals and suprasegmentals. Wang concludes that there are better learning outcomes when segmentals and suprasegmentals are both included in the instruction, and agrees with Zielinski (2015: 409) that research and teaching should focus on how both segmental and suprasegmental features combine to impact intelligibility and comprehensibility.

The review which I carried out for this section includes studies from a longer period of time (1974 - 2022) than Wang or McAndrews. I choose to focus on studies which concentrate on English (either as a target language or as the predominant L1), or on languages which include similar rhythmical structure to English (such as Dutch or German), and I include cross-linguistic studies. Several of the studies which I report on choose to focus on individual cues to suprasegmental features which have been shown to be teachable and learnable variables, and which contribute to intelligibility and comprehensibility - such as duration (Tajima et al, 1997) and F0 (Zetterholm & Abelin, 2019). I also include studies which not only focus on the narrower prosodic features of stress and intonation, but also those which



consider suprasegmentals in the broadest sense, such as pauses, phonological assimilation, and linking (Akita, 2005). McAndrews excluded such studies (McAndrews; 2019: 154). I also chose not to include only published articles, but also one PhD study (Capliez, 2016), because he focuses on French learners of LX English. Finally, I chose to include studies which used computer-assisted pronunciation training (CAPT) in their instructional programme (n=14). The use of technology is not recent, as we saw in section 5.6 with the use of Philippe Martin's F0 visualisation software (James, 1976). F0 visualisations are the most common form of CAPT used in the area of prosody-oriented pronunciation instruction, and remain popular in dedicated CAPT tools today, such as *Webpitcher* (Niebuhr, 2021), and *Blue Canoe* (Sakach, 2022). A table which lists all 32 studies can be found in the appendix to this report. It includes the following factors: focus of study, number and details of participants, target language, participants' L1, length of instruction period, pedagogical approach, presence of computer-assisted pronunciation training (CAPT), and the protocol and the method of assessment. All of the studies show an improvement in some or all of the features measured, with the exception of two (Kurt et al., 2014; Capliez, 2016).

Several studies on prosodic transfer have shown that prosodic features and other phonological features affect accentedness, intelligibility, and comprehensibility. For example, Okim Kang and her colleagues' (2016) study exploring these three constructs in Vietnamese-accented speech in English using listeners from a variety of linguistic backgrounds. They found that different groups of listeners used different cues, some reacting more to segmental features, and some reacting more to suprasegmental features. The language-specific nature of phonological transfer is not to be underestimated: whilst other aspects of a study, such as pedagogical methods used, number and age of participants, and so forth are all important, they are difficult to evaluate from a mere write-up in a scientific article. The linguistic context for a study, that is the target language and the language or languages of the participants is the most important independent variable of any study, and for this reason, I have grouped the studies by language in the following sections. Firstly, I will present some studies where the target language is English, followed by studies involving other target languages, before finally examining a few studies concerning English and French.

### 6.2.1. LX English

Firstly, we will consider intonation in pronunciation instruction studies. Myths as to the teachability or learnability of prosodic features still abound, for example Jennifer Jenkins' assertion that intonation is "unteachable" (2002: 99). The first studies we will review target exclusively, or mainly, the feature of intonation, beginning with two studies measuring learner perception. Judy Gilbert (1980) carried out

four small classroom studies using 36 Japanese and Cantonese speakers. Her first experiment concerned the identification of languages by using recordings stripped of their segmental data, the other three were on whether the participants could learn to recognise languages by intonation patterns with variations on the theme of a comparison of passive listening and active listening (“active listening” involves the students using a kazoo to mimic intonation patterns). Gilbert’s tests were all positive, and she concluded that people can differentiate between languages based on prosody alone, that training can increase learners’ ability to do this, and that mimicry has benefits (Gilbert, 1980: 116). Burcu Kurt and her colleagues (2014) attempted to demonstrate that instruction can aid learners to differentiate between ambiguous intonation patterns using inductive teaching with low intermediate learners from a variety of linguistic backgrounds. A statistical analysis of the results to this study was unable to confirm positive learning outcomes, although individuals appeared to have made gains, a result which the authors attribute to small group sizes (22 participants in total) (Kurt et al., 2014: 365). Several studies have shown the benefits on L2 English intonation production of using computer-assisted pronunciation training, by relying on different technologies to show learners the F0 curves of their pronunciation patterns (de Bot, 1983; Hardison, 2013; Zhuang, 2015). More recently, Maria Kostromitina and Okim Kang (2021) carried out a study using 75 Arabic- and Chinese-speaking English learners of three different proficiency levels, taking speech samples before and after a 15-week ESL course. They selected learning targets in segmental and suprasegmental using the functional load hypothesis and used a control group / study group to compare learning outcomes. Their results show no significant segmental gains across the levels, but clear gains in suprasegmental features, including speech rate for the advanced learners, and stress patterns for the intermediate and advanced learners. There were no significant gains in pitch range or the number of silent pauses, the other two features measured in this study. (Kostromitina & Kang, 2021: 11-12). Finally, Martha Pennington and Nick Ellis (2000) investigated recognition memory for prosody as distinct from lexis in English sentences. Their results showed that when advanced L1 Cantonese learners of English are encouraged to attend to intonational cues, they can, and do, use them to “interpret spoken sentences and remember their meaning accordingly.” (Pennington & Ellis, 2000: 386).

Many studies have investigated several prosodic features in English over the same period of instruction, for example Tracey Derwing and her colleagues’ (1998) study. They compared segmental, prosodic and general English instruction with 48 learners of mixed linguistic backgrounds split into three groups, and compared the results after 12 weeks of instruction for accent, comprehensibility, intelligibility and fluency. Their results are reported in table 6.3.1:

*Effects of instruction on performance in the three groups*

	NSP	Segmental	Global
Sentences			
Comprehensibility	no change	improved	improved
Accentedness	improved	improved <sup>a</sup>	improved
Narratives			
Comprehensibility	no change	no change	improved
Accentedness	no change	no change	no change
Fluency	no change	no change	improved

<sup>a</sup>significantly greater improvement than the global and NSP groups

**Table 6.3.1. The effects of different types of instruction on comprehensibility, accentedness and fluency over a 12-week ESL course. (Derwing et al., 1998: 403)**

Derwing and her colleagues use the evidence from this, the first controlled study of its sort conducted over an extended period, to support their view that pronunciation training should integrate prosodic and segmental features, and to call for more research with different approaches and in different contexts (Derwing et al., 1998: 408). The same team of three also published a smaller study the previous year which focuses on the acquisition of prosodic features in a smaller group of adult ESL learners who had been in Canada for an average of ten years (Derwing et al., 1997). The approach was to use Firth's "zoom" principal (Firth, 1992), which entails starting with broader prosodic features and to narrow in on smaller environmental features, and the results are very encouraging for any teachers working with adult learners, in that intelligibility improved over a 12-week course (Derwing et al., 1997: 229). In Mamiko Akita's (2005) study, also a 12-week course, results from a three-group controlled study showed that a prosody-oriented approach was effective in production and perception, and more effective than the segmental approach regarding production control study (Akita, 2005: 19). Yukie and Kazuya Saito (2017) found similar results with Japanese learners of English, but with a slightly shorter instruction time (three hours over ten weeks), with significant gains being measured in overall comprehensibility, word stress, rhythm, and intonation (Saito & Saito, 2017: 589). Mahmood Yenkimaleki (2019) took a different approach to the assessment of his participants, who were all trainee interpreters (Farsi – English) in his pre-test/post-test controlled study. Yenkimaleki used professional indicators of the effects to compare the effects of implicit and explicit instruction focusing on stress and intonation in a controlled study over a ten-hour course. Whereas there were no statistically different performances between the groups' TOEFL scores at the end of the course, independent raters found that the interpretation performance of the explicit instruction group had significantly improved between pre-and post-tests compared to the implicit instruction group and the control group.

Several studies using computer-assisted pronunciation training have also shown gains in multiple prosodic variables (Tanner & Landon, 2009; Hsieh et al., 2013; Lima, 2015). Mark Tanner and Melissa Landon's study is of particular interest because it involved self-directed learning with adult learners (17 to 54 years old) with a variety of linguistic backgrounds and lengths of stay in the US. Their pedagogical approach, cued pronunciation reading, involves some explicit instruction of prosodic forms, and self-directed practice. Over a 12-week course, results showed that the participants improved significantly on perception of pausing, perception of word stress, and controlled production of word stress, however there was no statistically significant gain in perceived comprehensibility (Tanner & Landon, 2009: 61). Oliver Niebuhr's (2021) study is very different from any other study in this review, partly because the training session lasted only one hour (guided instruction during a Zoom session with 60 diverse adult learners), but mainly because the tool he and his team developed is intended to improve prosody production with the aim of improving the perceived charisma of the learners. Anova tests show significant gains in six of the seven prosodic features measured, the exception being duration, which the tool (*Web Pitcher*) provides no direct visualisation or colour-coded feedback for (Niebuhr, 2021: 18). Darren Lascotte and Elaine Tarone (2022) use mirroring in their study, but as they used Ted Talks and language laboratories, I have counted this as a CAPT study. Part of the seven-week, 28-hour course involved explicit prosody instruction, and the learners mirroring Ted Talks having marked their own transcriptions of the talks for prosodic features. All the participants improved their intelligibility, and also their mean intensity and F0 levels between pre- and post-tests. Prosodic annotation is a technique I have been using in various projects since we developed a system based on Anthony Stenton's *SWANS* project (Stenton & Gimeno, 2011), and it is part of the protocol for the PIC (prosody, intelligibility and comprehensibility) project (Frost, 2021), which I will present in more detail in chapter seven.

Finally, concentrating on the contribution of stress and rhythm, both of Graeme Couper's studies (2006; 2022) found significant improvements, as well as interesting insights into learners' perceptions of prosodic features. Couper (2009) adopts a "cognitive phonology" approach, which he combines with quantitative data collection, and asks his learners to explain their responses using their own words. Graeme Couper visited Grenoble in 2018 and we met, along with several colleagues, to discuss the study he was designing on two-syllable word stress, and which was designed to be a replication study (Couper, 2022: 1), however unfortunately, at the time none of us had any suitable classes for running his protocol. Word stress is a particularly popular feature to explore in instruction studies, and several recent studies have shown measurable gains in short courses (Jung et al. 2017; Chen & Tian, 2020). My own research for my PhD study also showed significant gains in word stress perception after a short

instruction period (Frost, 2008): the instructional phase involved awareness raising through perception and production activities, but the majority of the work was production-based.

The studies on English prosody instruction which I have reviewed in this section all show gains, and often in short periods of instruction. My own research shows that prosody is often an area which LX learners – even English specialists – have usually not studied in depth, which means that there is often plenty of scope for improvement.

### *6.2.2. Other languages*

Most of the studies I have chosen to focus on concern LX English, but I would like to briefly mention two studies involving other languages. Federica Missaglia (1999) compared the effects of a prosody-centered approach to a segmental approach for low level L1 Italian learners of German, the results of which were rated by L1 German speakers. This study is interesting to me because although Italian and German are both stress languages, Italian is a Romance language, as is French, and German and English have similar prosody (Kember et al., 2017: 3188). Missaglia's results show greater positive effects on global pronunciation ratings for the suprasegmental group than the segmental group (Missaglia, 1999: 553), and she notes that "accentuation and intonation have a controlling function over syllables and segments". (Missaglia, 1999: 551). The other study which I would like to mention concerns bilingual Catalan/Spanish learners of French, and it is of particular interest because the authors, Florence Baills, Charlotte Alazard-Guiu, and Pilar Prieto (2022) adopt an embodied approach to teaching pronunciation, which targets prosody in particular. They used iconic hand gestures to accompany intonation patterns and accentual phrases over three sessions between pre-test video recordings and the post-test video recordings in a controlled study, and the learner recordings were assessed by L1 French experts for comprehensibility, fluency, accentedness, segmental accuracy and suprasegmental accuracy. Results showed a significant increase in all areas, with higher effect sizes in all areas for the group which received the gestural training (Baills et al., 2022: 795).

### *6.2.3. The French context*

The two languages which are the most relevant to this report are English and French - or rather English as an LX for L1 French learners - so we will briefly examine three studies which investigate the acquisition of French prosodic structures by English L1 learners. Both E.F. James (1976) and Debra Hardison (2004) used F0 visualisation tools to aid L1 English learners to improve their intonation in French. In James' three-group controlled study, he found that visualisation helped, but the auditory

feedback had no significant effect. Regarding the rater reviews however, James noted that even when articulation was poor, an utterance with correct intonation was very acceptable, whereas utterances where articulation was good but intonation is poor were much less acceptable (James, 1976: 242). Debra Hardison (2004) used a similar instruction method, and compared “filtered” production samples (i.e. prosody only) and unfiltered production samples, and she found improvement both at the segmental and suprasegmental levels. Cécile Chapagne-Muzar and her colleagues (1993) conducted a controlled study, where instruction focused on perception then production. After a 12-lesson course, they obtained a gain in the production of the experimental group, measured by L1 French listeners rating segments, intonation, rhythm, and global impression on 5-point scales. Also, in a similar way to Couper (2006; 2009; 2022), Hardison asked her learners for qualitative feedback on the learning process, and many of the comments were very positive and give insight into learner perceptions of the usefulness of prosody-oriented instruction (Hardison, 2004: 46). Finally, Charlotte Alazard and her colleagues, Corinne Astésano and Michel Billières (2010) present a pilot study examining the effectiveness of the use of the Verbotonal method on the reading aloud abilities of four L1 English speakers, and the perceptual test showed a marked improvement in the lower ability learner who received the Verbotonal instruction, a result which was confirmed in her PhD study (Alazard, 2013).

The last studies we will examine are of L1 French learners, starting with another pilot study, this time carried out shortly after the publication of Chomsky And Halle’s *Sound Patterns of English* (1968). Marc Schnitzer (1974) – apparently in a single session – explicitly taught stress placement rules and laxing/tensing rules to three learners, and found that it greatly helped them with their realisation of word stress on real and nonsense words, up to 900% in the case of one learner (Schnitzer, 1974: 294). Although this study has obvious limitations, it is indicative of the effect that Chomsky and Halle’s work had on pronunciation teaching and research in the years following its publication. Kees de Bot and Kate Mailfert (1982) conducted a study with a group of adult L1 French learners of English which confirmed de Bot’s earlier findings teaching English to Dutch learners, which were actually published the following year (de Bot, 1983). The methodology included explicit prosody-based instruction using stripped recordings and graphic representations to help understand and reproduce desired prosodic patterns. Both the learners in the Dutch study and the French study showed “remarkable progress” (de Bot & Mailfert, 1982: 73). As in several studies presented in this section, the researchers asked learners for feedback on their learning experience, and the comments are very useful, both the negative comments (some students disliked passive listening to electronically modified recordings) and the positive comments about the perceived usefulness of an increased awareness of the differences between French and American English intonation (de Bot & Mailfert, 1982: 76-77). Marc Capliez (2016) carried out a series of production and perception experiments in a three-group study (prosody-based

instruction, segmental instruction and a control group) on L1 French university students learning English with levels ranging from A2 – B2. Capliez’s pedagogical approach during the ten 40-minute sessions was very teacher-centered, involving much choral and individual repetition. Unfortunately, Capliez failed to find any significant differences between the groups in his study, suggesting that perhaps small sample size was a factor (Capliez, 2016: 179). The last study I would like to mention in this section is particularly interesting because it investigates L1 French learners of English, and L1 English learners of French in a replication study of Emmanuel Dupoux and his colleagues’ study on “stress deafness” in L1 French learners of Spanish learners (Dupoux et al., 1997), but with the addition of prosody instruction. The instruction used perceptual fading, i.e. playing speech recordings with modified, exaggerated acoustic correlates, in this case duration, which is gradually decreased over time. The trained French group performed significantly better than the untrained group, and this, taken with the other findings of the study, shows that this method can be an effective way to improve perception of stress in English by L1 French listeners (Carpenter, 2015: 104).

The majority of studies focusing on French learners confirm the findings of the studies in the previous sections: the learning outcomes of prosody-based instruction, even with adult learners, can be considerable.

### 6.3. Deconstructing intelligibility: towards a model

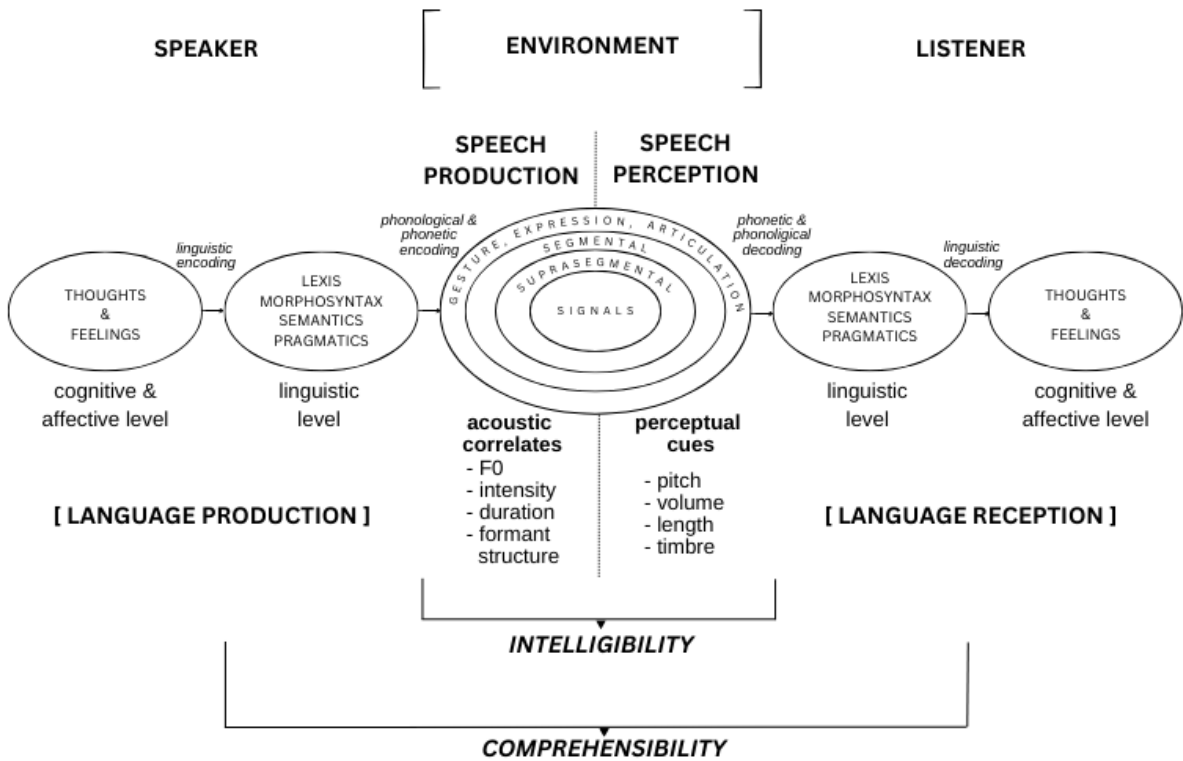
In the nearly 30 years since Munro and Derwing’s (1995) study on accent, intelligibility and comprehensibility, the community of teachers and researchers in the field of pronunciation teaching research in North America, and increasingly throughout the world, has aligned itself with the same definitions of these constructs. However, in their seminal article, Munro and Derwing essentially use the term “comprehensibility” as a synonym for intelligibility, differentiating instead between objective and subjective measures for the two constructs. As Fayer and Krasinski point out: “Subjective designs involve evaluative judgments and some type of rating scale, while objective tests use operative interpretation of the deviant message, such as rewriting or rephrasing” (Fayer & Krasinski, 1987: 314). Foote & Thomson (2021) suggest that speech pathologists and other medical professionals lack a solid grounding in the field of pronunciation instruction, and that there is much that speech pathologists can learn from linguists and LX pronunciation experts (2021: 20-21). I would suggest that we too can learn much from clinical professionals’ approaches to speech disorders. This notion of objective or subjective assessment of these constructs is something we see in the clinical community, as we saw in section 6.1.2. Nicholas Pommée refers to “perceptual” and “objective” assessment methods of intelligibility and comprehensibility (Pommée et al., 2022: 26). Furthermore, the clinical community

clearly hold intelligibility and comprehensibility to be two different constructs, which need measuring in different ways: intelligibility, or a lack thereof, is a result of acoustic parameters in the speech signal and the effect which that produces on a listener, whereas comprehensibility involves many more linguistic factors, such as lexical and morpho-syntactic control (Yorkston et al., 1996: 55; Pommée 2022: 11). This distinction is now increasingly being understood in the pronunciation and teaching research community (Isaacs and Trofimovich, 2012; Trofimovich and Isaacs, 2012), and yet both comprehensibility and intelligibility are repeatedly used as measures in pronunciation studies, using the same definitions: “*intelligibility* (the degree to which a listener understands a speaker’s intended message), *comprehensibility* (the degree of effort required for a listener to understand L2 speech) and *accentedness* (the degree of difference from an expected accent)” (Levis, 2022). These definitions of intelligibility and comprehensibility clearly conflate two separate levels: the level of defining the constructs (phonological considerations, or wider linguistic considerations), and the level of measuring the constructs (subjective or objective measures). In Talia Isaacs and Ron Thomson’s most recent paper (2022: 2), the authors still consider perceptual (subjective) constructs to include “comprehensibility, accentedness and fluency” (Isaacs and Thomson, 2022: 125), rather than seeing both comprehensibility *and* intelligibility as perceptual constructs, but linked to different features, measurable in different ways. So, in fact, successive studies have measured the construct of intelligibility *only* using objective measures, and measured comprehensibility *only* using subjective measures. My research using the Common European Framework (Frost & O’Donnell, 2013; 2015) and developing the descriptors (Frost & O’Donnell, 2018) which I will present in the next chapter have led me to two conclusions. Firstly, the construct that is being measured in pronunciation instruction studies with raters is not “comprehensibility”, but “intelligibility”, however it is being rated in a subjective, not in an objective manner. Secondly, all assessment of oral production is necessarily subjective, since it involves the judgment of other human beings; however, that judgement can be given a certain degree of objectivity through the use of tools such as calibrated descriptors, or by acoustic measurements, the values of which have been pre-defined. In a clinical context, the objective measures of intelligibility involve biological and behavioural tests relating to pathologies such as aphasia, or stroke-related speech difficulties, whereas in language teaching, the area of interest is learner interlanguage. In both contexts, intelligibility must be deconstructed into various components which can be “treated”, i.e. improved, and often clinical approaches involve learning and practice, as does pronunciation instruction.

Based on the models of speech perception and production which I presented in section 3.5, figure 6.4 below represents the process of speech production and perception in order to illustrate the constructs of intelligibility and comprehensibility: what they refer to, and what we can measure in order to



understand them. The process of speech communication is represented as a movement from left to right, and separated by a vertical dotted line in the middle of the diagram, which represents the physical interface between speaker and listener, i.e. the physical environment of time and space in which both speaker and listener interact. The thoughts and feelings of a speaker pass through stages of linguistic encoding, ending with phonological and phonetic encoding, resulting in the activation of motor neurons and the production of an acoustic signal (Levelt, 1989; 1999). The **intelligibility zone**, or the area where intelligibility can be measured, exists only after phonological and phonetic encoding has happened, and once the speaker has actually produced an auditory speech signal and the accompanying visual signals. Once the speech signal has been received by a listener and compared to phonetic, phonological and syntactic information based on linguistic knowledge (Lieberman & Mattingly, 1985: 9), we are no longer in the domain of intelligibility, as other linguistic knowledge has been activated. This linguistic knowledge is not only present on the part of the listener, but may also involve top-down compensatory techniques operated by the speaker to compensate for impaired language (Pommée et al., 2022: 11). The **comprehensibility zone** therefore encompasses the intelligibility zone, and goes beyond it into the areas of linguistic encoding and decoding.



**Figure 6.4. An embodied cognition model of speech**

In figure 6.4, the concentric circles in the centre of the model represent linguistic signs: I use the term “signals” at the centre, to represent signs at a more abstract level. At the perceptual level, access to this deep level is achieved by the processing of the visual and/or auditory cues of the surface

phonological features. At the level of production, the signals become concrete auditory and/or visual signs as they give rise to prosodic features, other suprasegmental features (such as pauses), segmental features, and visual cues to speech (gestures, facial expressions, and the visible movements of articulators). I chose to place “suprasegmental” inside “segmental” because of the impact prosody can have on segments: this is particularly true of a language such as English, where the presence or absence of stress often determines the perceptibility of phonemes, particularly vowels. Intelligibility and comprehensibility are both dependent on segmental *and* on suprasegmental features (Kang et al., 2021: 32), and as we have seen, these features are composed both of measurable physical realities, or correlates, and the perceptual cues they represent for the listener. However, I consider the construct of comprehensibility to be of limited use to pronunciation studies, as it depends on many factors other than pronunciation. In addition, I take issue with the definitions generally used to define the constructs of both intelligibility and comprehensibility, and I think it would be far more useful for both teaching and research to use definitions more in line with those used by speech pathology researchers.

In this chapter, I have presented the constructs of accent, intelligibility and comprehensibility, both in terms of applied linguistics studies and studies examining learning outcomes. We have also seen the clinical approach to defining intelligibility and comprehensibility, and that those two constructs are considered to be very different in that community, and that they can be measured in both perceptual, subjective ways, or in more objective ways. Regardless of how we define these constructs, the majority of research on prosody demonstrates three things: firstly, that prosody has an impact on how successfully spoken language is perceived as being easily understandable, and on how well it is actually understood. Secondly, prosody can be taught and learned effectively, even in short periods of time. And thirdly, that improved prosody leads to improved understanding and perceived ease of understanding. Most of the studies reviewed in this chapter have isolated prosodic features or cues to prosodic features, in order to measure different effects, but the overall picture from the research is that an integrated approach is the best way to improve intelligibility, where prosody is not overlooked or taught as a separate skill, but as working together with segmental features in an interdependent system (Derwing, et al., 1998; Zielinsky, 2015; McGregor & Reed, 2018). Such an integrated approach must account for the speaker and the listener, the cognitive and physical factors involved in pronunciation, and the auditory and visual cues to prosody. And in order to understand which features to concentrate on, and which learning objectives to set, the environmental factors pertaining to each individual learner in their use of the target language must be considered. In the next and final chapter of this report, I will describe some ideas which can help learners, teachers and researchers to set and achieve appropriate pronunciation learning goals.

## CHAPTER SEVEN

### Prosody in LX learning, teaching and assessment

Having explored prosody in research from the fields of phonetics, phonology, psycholinguistics, and other sciences, in this final chapter, I would like to concentrate on learning and teaching. My chief motivation for the research that I conduct is how it will benefit learners, who are mainly – but not all – L1 French speakers. I have chosen to focus on pronunciation to improve learners’ intelligibility, but also to help them to perceive and understand spoken English. Firstly, I will explore some issues concerning norms and models for pronunciation instruction. Secondly, I would like to see how we can assess pronunciation and set learning objectives. Thirdly, I will develop some ideas based on embodiment and foregrounding prosody in pronunciation instruction, including pedagogical principles and content inside and outside the classroom. Next, I will explore some of the ethical issues surrounding pronunciation instruction, before finally presenting a few ongoing projects which I am currently involved in and future directions for my research.

#### 7.1. Norms, varieties and models<sup>14</sup>

Pronunciation teaching is where the science of phonetics has its roots, and it is still an important field of both theoretical and applied research today. When *Dhi Fonètik Tîcerz’ Asósiécon* was founded in 1886 (later changing its name to the *International Phonetic Association*), it aimed to help younger learners to acquire “a realistic pronunciation of foreign languages” (International Phonetic Association, 1999: 194). Traditionally, language teaching, and in particular teaching pronunciation, was very prescriptive, and the focus was generally on accent reduction, or accent modification, in order to achieve a “correct” pronunciation which would be identical to that of an L1 user (Pedrazzini, 2016), specifically the pronunciation of a variety considered prestigious, such as received pronunciation (RP) for English.

When discussing models of varieties of language in the context of pronunciation teaching, it is important to remember that any model is merely a simplified representation of a complex reality, and

---

<sup>14</sup> In this section, “models” does not refer to models about processes such as language perception, production or learning as previously discussed, but to representations of varieties of languages which are often used for teaching towards, such as General American, or Received Pronunciation.

this is true of received pronunciation or General American, or any model of any language. As Peter Roach (2009) explains:

Each individual's speech is different from any other's; it follows from this that no one speaker can be taken to represent a particular accent or dialect, and it also follows that the idea of a standard pronunciation is a convenient fiction, not a scientific fact. (Roach, 2009: 163)

And yet, the question of which model teachers should teach towards is still a relevant one, not least for pragmatic reasons such as the development of teaching resources, and for high stakes language certifications, such as those necessary for citizenship status or for university entry.

### 7.1.1. *The "native speaker debate"*

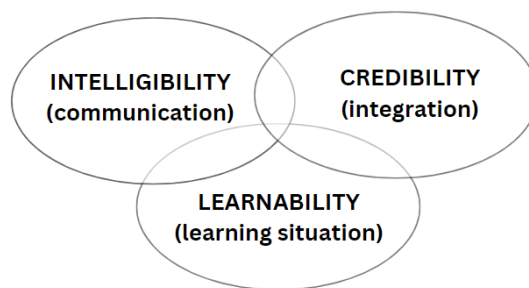
At the heart of this debate, as it was framed by John Levis (2005; 2020), is the fact that language teaching professionals have always had a choice between "the nativeness principle", which involves teaching towards a native speaker model, and "the intelligibility principle". This is not a new debate by any means, and was evident in the contrasting language pedagogies of the late 19<sup>th</sup> century, which opposed on the one hand accuracy and explicit instruction (the reform movement), and on the other hand communication-based L1-influenced instruction (the direct method) (Richards and Rodgers, 2001: 11). This form of "standard language ideology" (Lippi-Green, 1994) can be viewed as a kind of gate-keeping, or even "a manipulative attempt to improve learner behaviour" (Holliday, 2006: 387). Regarding learner preferences as to teachers' L1, as we saw in section 3.3, studies have produced varying findings: some studies have shown that learners prefer "native speakers" (Hertel & Sunderman, 2009), whereas other studies have shown that learners do not always prefer "native speakers" (Llurda, 2004).

More recently, another norm has been added to the debate of LX English-teaching, that of the English Lingua Franca core (Jenkins, 2000; 2002), which I presented in section 5.1. The list of phonological features (table 5.1) suggested by Jenkins (2002: 99) is intended as a set of pronunciation learning goals which she opposes to traditional "native speaker goals". However, many of Jenkins' conclusions were based on hours of recordings between Asian users of English, and while certain consonants and consonant clusters undoubtedly pose major problems for users of many Asian languages, each context is different, and it is difficult to imagine how a single model of common features could correspond to all the learners in a single classroom, let alone throughout the world.

### 7.1.2. *Aims for LX English pronunciation*

In terms of reception skills, the varieties of language to which learners must be exposed will depend on their needs: for most learners, the more varieties the better, both of L1 Englishes and LX Englishes. In terms of production, implicit within the native speaker debate is the question of what the aim for a language learner should be: should it be to lose their accent, or to modify it in such a way that their personal and/or professional goals may be achieved? Some learners may want to aim for, and even achieve near-native pronunciation skills (Bongaerts et al., 2000; Timmis, 2002; Derwing, 2003), while others may want to speak with an accent to take advantage of its positive effects (Derwing and Munro, 2008: 484). The goal of “native-speaker like pronunciation” may be appropriate, for example, for a teacher aiming for a career in language teaching: as we saw in section 6.3.3, the teacher recruitment process for English teachers in France now accepts a variety of regional L1 accents in a candidate’s pronunciation, but a candidate must also have detailed knowledge of only two models (both in terms of reception and production), GA and SSBE – which they must be able to demonstrate themselves – in order to pass muster.

This debate over the goals of pronunciation instruction at the level of production is often framed in terms of “**accent modification**” or “**accent addition**” (Thomson, 2013; 2014). The term “**accent reduction**” is a little different, and is often used as a selling point for pronunciation courses with unrealistic and unverifiable claims, such as those aimed at immigrants hoping that reducing their accent will help them to integrate in their new country (Thomson, 2013: 1-2). Accent modification is perhaps a more appropriate term for learners with intelligibility as their primary aim. Joanne Kenworth (1987:3) stated that although “most people now think that [native speaker-like pronunciation] would be an inappropriate goal for most learners”, she mentioned that for certain professions - such air traffic controller - the goal must be a pronunciation which can be clearly understood, even in difficult conditions. Under such circumstances, accent modification would be the most appropriate goal, and the model or models used as example for teaching are not as important as the pronunciation issues of the individual concerned. Any pronunciation assessment must take into consideration learners’ pronunciation at the beginning of instruction, and establish goals balancing learner wishes and needs with the constraints and affordances of the learning situation. In this sense, negotiating and renegotiating goals for pronunciation instruction amounts to constantly balancing the goals of intelligibility and credibility, and weighing them up against what is learnable, as figure 7.1.2 represents:



**Figure 7.1.2. Issues for negotiated LX English pronunciation learning goals (production)**

Pronunciation instruction is usually part of a language course which will also include vocabulary and grammar objectives, and work on other linguistic and non-linguistic skills, and so time given to pronunciation must be weighed against other considerations. There are also many individual and environmental factors which influence what is learnable in any given situation, as we have seen in previous chapters. In some contexts, accent modification may be a realisable goal, but for many shorter courses, it is simply unrealistic.

### *7.1.3. Accent addition or accent exposure (pronunciation for reception)*

The term “accent addition” is often used in the context of International English(es) by applied linguists and teachers who advocate teaching core features such as the ELF core (Jenkins, 2002). As Jenkins explains; “‘accent addition’ is being promoted in accordance with the goals of additive bilingualism and in tune with the current emphasis on learner choice (Jenkins, 2004: 115). The term implies that learners are capable of “switching” between languages according to situational requirements, while retaining their own language and identity, a concept which chimes with many language teachers. Olle Kjellin, for example (1999) uses the term accent addition: he uses an embodied and prosody-centered approach when teaching Swedish to immigrant adult learners who need to be able to integrate rapidly into a country with a language to which they have had little or no exposure prior to arrival. I prefer to think of the term “accent addition” in terms of the addition of targeted pronunciation features to a learner’s interlanguage, rather than the addition of a new accent, and I would relate this process to **accent exposure**, or what Richard Cauldwell refers to as teaching phonology for listening (Cauldwell, 2013). In the motor theory of perception (sections 2.2.1 and 3.5), speech perception triggers an active response in the listener, whose speech centres actively (if not actually) produce a similar section of speech to that under analysis. It follows that if the listener is not familiar with the sounds of the language in question and cannot themselves produce those sounds, then the process will not be completed. If, however, a listener is able to produce the sounds under analysis, there is a greater chance that they will be able to successfully perceive and segment spoken discourse into manageable chunks, before moving on to higher level cognitive operations, such as linguistic analysis, comparison to the items in the mental lexicon, and ultimately, comprehension. By exposing learners to

phonological features which are not yet part of their own interlanguage pronunciation, and by practising these features both in terms of reception and production, the aim is not necessarily to permanently integrate these features into learners' own pronunciation patterns: perhaps they may be able to imitate the features in isolation in the context of a language classroom, and only realise them partially in spontaneous speech outside the classroom context. In this way, even if the features are not a permanent part of a learner's idiolect, they would be more likely to be able to understand another speaker who uses such features. Rather than teaching towards *one* model, a model which as we have seen would by definition not constitute a reality, I would argue that it makes more sense to expose learners wherever possible to a variety of real accents, both international accents, as John Murphy (2014) suggests, or regional English accents, as Kizzi Edensor (2008) advocates, and where a variety or certain varieties are more likely to be encountered by learners, then those varieties should be given priority.

#### 7.1.4. *Listener training*

Another approach to the question of intelligibility and comprehension is to train listeners, i.e. not the actual LX learners, but the set of individuals who will have to listen to foreign accented speech and understand it. Listener training has been shown to improve LX comprehension (Clarke and Garret, 2004), and exposure to regional accents has been shown to improve the chances of LX English learners understanding those accents (Edensor, 2008). There are also programmes in Swedish and American universities to train students to listen to accented speech (Jeong et al., 2021) which the authors of these programmes believe will improve comprehension of world Englishes, and also challenge students to accept a wider range of variations (Jeong et al., 2021: 143). While the goals of challenging standard language ideology and native speakersism within the relative safety of a phonetics class or a language course in a Swedish or American university are perfectly laudable, it is of course impossible to train all the people who will have to listen to accented English in all the professional and personal contexts where learners will have to make themselves understood. However, accent exposure from an early age in language classes in schools, using a variety of L1 and LX accents rather than perpetuating teaching through the use of abstract models would be an effective way to challenge native speakerism, and to improve communication when learners go on to use English in later life.

## 7.2. Prosody & assessment

It may seem counter-intuitive to include a section on assessment before developing ideas about teaching – after all, usually learning precedes assessment. However, I would suggest that it is essential to begin any instructional programme with an assessment of the level of the learners, and that this is part of the necessary initial step of a needs analysis. Assessment can serve many purposes, the three most-often cited being diagnostic, formative and summative. Alderson (2005: 6) states that less attention is paid to diagnostic assessment compared to high stakes tests, such as those required for citizenship or employment. Bachman & Palmer (1996: 98) state that “[d]iagnosis involves identifying specific areas of strength or weakness in language ability so as to assign students to specific courses or learning activities”, and this is very much the approach which I feel is useful in the context of pronunciation instruction, since learners generally have very different pronunciation issues, even if a placement test has been carried out and learners are streamed into groups by level, as was the case in the IUT (*Institut universitaire technologique*) where I worked for the first ten years of my teaching career in the Grenoble. IUTs in France have a national curriculum for each department, and for computer science, the programme at that time (2003) did include some detail on pronunciation: “*Principes élémentaires de phonologie: prononciation des sons, schémas intonatifs, accents de mots, accentuation de phrase.*” (basic elements of phonology: the pronunciation of sounds, intonation patterns, word stress and sentence stress) (Ministère de l’Éducation Nationale, de la Jeunesse et des Sports, 2003). There was however, no nationally agreed test for English in the IUT system, and so the APLIUT (*Association des professeurs de langues des IUT*), of which I was an active board member, decided to draft a motion announcing our position to support and integrate the use of the Common European Framework of Reference for Languages (CEFR). I had been using the CEFR since it was published two years earlier, and like many of my colleagues, I found it to be an extremely useful tool, especially for teaching pronunciation towards intelligibility, which was very much the pragmatic approach adopted by my most of my colleagues teaching English for specific and academic purposes in the APLIUT.

As we saw in the first section of this chapter, pronunciation instruction aims may be more towards a native speaker model (or other models, such as Jenkins’ *Lingua Franca Core*), or towards intelligibility. In the previous chapter, I attempted to show three things: firstly, that intelligibility was principally the result of pronunciation, whereas comprehensibility was the result of pronunciation *and* a range of other linguistic factors, such as morpho-syntax, semantics, and pragmatics; secondly, that prosody was very often an important factor in intelligibility; and thirdly, that both intelligibility and comprehensibility could be measured in ways which were either more subjective, or more objective.



In Isaacs and Trofimovich’s most recent publication (2022), the authors list comprehensibility as an example of a “perceptual construct” (Isaacs & Trofimovich, 2022: 125), but do not include intelligibility in this list. By considering “raters’ accounts of the linguistic features they reportedly pay attention to when scoring L2 speech” (Isaacs & Trofimovich, 2022: 147), the authors are moving closer towards an approach where measures of accentedness could be used to assess what they continue to refer to as comprehensibility (the word “intelligibility” is not mentioned once in the chapter). In my own research, the deconstruction of factors affecting accentedness, and therefore intelligibility, has been my position and my approach to pronunciation assessment since the ELLO project (Frost & O’Donnell, 2013; 2015), and which led me to compile the prosody descriptors (Frost & O’Donnell, 2018; Frost 2021) which I will present in this section. While I was carrying out the longitudinal study on the oral production and oral interaction of three successive cohorts in the applied foreign languages (LEA) department at Savoie Mont Blanc University, I spent many hours with my collaborator, Jean O’Donnell, poring over the scales that we used for assessing the students. These were the scales as they had been presented by our colleagues as part of the WebCEF international project on collaborative oral assessment (Baten et al., 2013). The descriptors which we used were slightly adapted versions of the oral production descriptors from the CEFR, and used the following criteria: accuracy, range, fluency and coherence (Council of Europe, 2001). These criteria were adequate for the majority of our research, and I was able to train the first-year students in their use, so that they too could use them to rate their performance and the performance of their peers accurately – in fact, half a grade more severely than we did (Frost & O’Donnell, 2013; 2015). We had chosen not to add phonological control to the project, because we both taught pronunciation classes, and found the phonological descriptors too general to be of use: in this regard, the CEFR is both successful and frustrating because the descriptors are *common* to all European languages. The original descriptors for phonological control are pictured in table 7.2a:

<b>PHONOLOGICAL CONTROL</b>	
<b>C2</b>	<i>No descriptor available</i>
<b>C1</b>	<i>Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.</i>
<b>B2</b>	<i>Has a clear, natural, pronunciation and intonation.</i>
<b>B1</b>	<i>Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.</i>
<b>A2</b>	<i>Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.</i>
<b>A1</b>	<i>Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.</i>

**Table 7.2a. Common European Framework of Reference for Languages “phonological control descriptors” (Council of Europe, 2001: 28)**

The lack of detail relevant to the precise features of a given L1 and of a given target language is one reason that some raters find that they are not relevant, for example to questions of stress-timing and chunking in English (Harding, 2016: 26). In 2013, we presented the ELLO project (Frost & O'Donnell, 2013) at a conference in Antwerp entitled "Language teaching in Europe: Time for a new Framework?" with the participation of members of the team that had written the CEFR. During the round table discussion, several people mentioned their criticisms of the CEFR descriptors, and the answer was always "then make your own", or words to that effect. So, this is what I did, writing the "prosody descriptors" (table 7.2b) and together we went through the long and painstaking process of calibrating the descriptors, pegging them to the CEFR levels for average French L1 learners of LX English, using the recordings we had obtained and graded during the ELLO project (Frost & O'Donnell 2018). The descriptors were part of the PIC pilot study (Frost, 2021) and continue to be used in the project.

PROSODY DESCRIPTORS: for self-assessment / teacher assessment / peer assessment of students' oral production in English

Level	RHYTHM & STRESS	SOUNDS			INTONATION
	<u>Is the correct syllable stressed &amp; marked correctly?</u>  (word stress & focus, i.e. nuclear and contrastive stress)  <u>The stressed syllable should be:</u>  • higher (there is some pitch change) • louder • longer • pronounced more clearly / "correctly"	Reduced syllables  <u>Reduced syllables are usually:</u>  • less high • less loud • shorter • pronounced less clearly / "correctly"  E.g. schwa /ə/, (doctor); /ɪ/, short final /i/, etc. (happy); syllabic /n/ & /l/ (button, bottle). 30-35% of all vowels should be schwa!	Stressed & unreduced vowels  • Are the stressed & unreduced vowels closer to English or closer to the speaker's native language?	Connected speech (phonotactics)  • Contractions ('ll, it's, gonna etc.) • Linking (e.g. an egg) • Linking with /j/ & /w/ (e.g. go-/w/-away) • Deletion of final /t/ & /d/, etc., (e.g. first question) • Assimilation (e.g. Breaking Bad > Breakimbad) • Geminates (e.g. keep playing) • Etc.  Note: Higher level speakers will pause between sentences and tone units. Lower level speakers hesitate more often, disrupting the flow.	Intonation  • Are the intonation patterns varied and appropriate to the speaker's intentions? • Is the range clearly marked or is the intonation flat?  Note 1: Monologues may not provide opportunities for different patterns. Note 2: Some younger native speakers use almost exclusively HRT (High rising Terminal). This is also true of learners exposed to lots of recent English, including film & TV series.
C2	Can place and mark word stress & nuclear stress at will using all 4 cues without disrupting flow. Errors are extremely rare.	Can reduce the full range of forms, including syllabic /n/ & /l/, etc. resulting in a natural-sounding alternation of strong & weak syllables.	Can produce all / nearly all vowels in a stable and consistent accent close to e.g. US, GB.	Can produce native speaker level connected speech phenomena making for a smooth and natural sounding flow. Can use more or fewer connected speech phenomena according to register, speed, etc.	Can produce native level & natural-sounding intonation patterns appropriately, including for attitudes, emotions, humour, etc.
C1	Can almost always correctly place word stress and nuclear stress. Can clearly mark stress with all 4 cues. Errors are rare.	Can reduce nearly the full range of forms. Strong & weak syllable alternation is evident most of the time.	Can produce nearly the full range of English vowels with a stable accent often close to e.g. US, GB.	Can produce mainly smooth and natural-sounding connected speech. Nearly the full range of phenomena, etc., but the occasional missed opportunity, especially in longer sentences. Hesitations are rare.	Can produce nearly all natural-sounding intonation patterns appropriately, even for attitudes, emotions, humour, etc.
B2	Can correctly place both word stress and nuclear stress nearly all of the time using all 4 cues to varying degrees.	Can partially or fully reduce most possible reductions. Strong & weak alternation is often evident.	Can produce nearly all vowels, but L1 transfer is noticeable on many vowels.	Can produce a variety of phenomena, including assimilation & deletion well over half of the time, especially in shorter sentences without many hesitations. Can produce fluent stretches of less "staccato-sounding" speech with lots of contractions.	Can produce a wide variety of appropriate patterns, including some attitudes, emotions, humour, etc. Can often produce a good range between high and low tones.
B1	Can use the 4 cues but not consistently. Can correctly place word stress most of the time. Can correctly place nuclear stress most of the time, especially in shorter sentences.	Can partially reduce half of all reduced syllables. Some full reductions, especially schwa. Strong & weak alternation is evident on shorter sentences.	Can produce most vowels but L1 interference is evident on most vowels, but rarely causes comprehension problems.	Can produce about half of all possible linking phenomena, including deleted /t/ and /d/, assimilations, etc., but hesitations are fairly frequent in longer sentences.	Can produce a variety of appropriate patterns, including the more obvious attitudes, emotions, etc., Can sometimes produce higher tones.
A2	Can usually place word stress correctly. Nuclear stress is placed correctly some of the time. Occasionally uses all 4 cues together.	Can produce a few partially reduced syllables, mainly schwas.	Can differentiate between tense & lax vowels & diphthongs, but L1 interference very evident on all / most vowels.	Can occasionally produce contractions ('ll, 'd, gonna, wanna, etc.) Can link between words less than half the time with word-final consonants and some /j/ & /w/ between words (e.g. go out). Frequent hesitations.	Can produce some appropriate patterns, especially on shorter sentences. Range between low and high is minimal.
A1	Can audibly place stress on isolated words or short sentences only. Limited control of the cues which mark stress so the stressed syllable is usually difficult for the listener to identify.	Can rarely if ever reduce syllables; reduced syllables are usually pronounced the same as stressed syllables.	Can produce a limited number of English vowels on isolated words & expressions. L1 vowels in place of target vowels on almost all / all vowels.	Can produce basic and isolated contractions (e.g. I'm, it's, gonna, etc. Can produce word internal /j/ & /w/ linking (going). Occasional linking between words when a word-final consonant is followed by an initial vowel (e.g. An-egg). Very frequent & long hesitations.	Can occasionally produce appropriate intonation patterns on short learnt phrases (e.g. greetings). Range between low and high is minimal.

Table 7.2b. The prosody descriptors (Frost & O'Donnell, 2018)

The prosody descriptors have a dual objective: firstly, they are intended for use by teachers and/or students to assess the level in each of the categories, and secondly, to fix learning objectives for pronunciation instruction. The order of presentation of the three categories (rhythm and stress, sounds, and intonation) corresponds to the relative importance of those features to the perceived intelligibility of French-accented speech in English. The category “sounds” is separated into three sub-categories (reduced syllables, stressed and unreduced vowels, and connected speech phenomena) for the same reason. The phonemes of English are therefore analysed via the effect that the stress and rhythm of English imposes on them, which encourages both segmental and suprasegmental features to be viewed as part of a single, integrated system (Derwing, et al., 1998; Zielinsky, 2015; McGregor & Reed, 2018).

Following the publication of the prosody descriptors, I was contacted by Brian North in 2016, who was in the process of writing the new descriptors for the *Companion Volume* to the CEFR (North, 2020; Council of Europe, 2020). The *Companion Volume* updates but does not replace the CEFR descriptors, which remain valid (Council of Europe, 2020: 3). The only descriptors which have been revised are the phonological control descriptors which were, as I mentioned in the previous section, so general as to be of limited pedagogical value. The primary objective of the CEFR is not, as many people believe, to assess learners’ language abilities, and The Council of Europe is not the same as, or part of, the European Union – it is Europe’s most important human rights body. As such, the CEFR and its descriptors aim to improve communication between the citizens of Europe and their institutions (technical alignment), and to improve language teaching and learning by promoting an action-oriented approach to language (North, 2020: 550). I was asked to participate in the writing of the new phonological control descriptors in the *Companion Volume* (table 7.2c), and I was also one of more than 1500 people who helped in the validation process of the new descriptors (mainly for different types of mediation, such as cultural and technological mediation).

	Phonological control		
	Overall phonological control	Sound articulation	Prosodic features
C2	Can employ the full range of phonological features in the target language with a high level of control – including prosodic features such as word and sentence stress, rhythm and intonation – so that the finer points of their message are clear and precise. Intelligibility and effective conveyance and enhancement of meaning are not affected in any way by features of accent that may be retained from other language(s).	Can articulate virtually all the sounds of the target language with clarity and precision.	Can exploit prosodic features (e.g. stress, rhythm and intonation) appropriately and effectively in order to convey finer shades of meaning (e.g. to differentiate and emphasise).
C1	Can employ the full range of phonological features in the target language with sufficient control to ensure intelligibility throughout. Can articulate virtually all the sounds of the target language; some features of accent(s) retained from other language(s) may be noticeable, but they do not affect intelligibility.	Can articulate virtually all the sounds of the target language with a high degree of control. They can usually self-correct if they noticeably mispronounce a sound.	Can produce smooth, intelligible spoken discourse with only occasional lapses in control of stress, rhythm and/or intonation, which do not affect intelligibility or effectiveness.  Can vary intonation and place stress correctly in order to express precisely what they mean to say.
B2	Can generally use appropriate intonation, place stress correctly and articulate individual sounds clearly; accent tends to be influenced by the other language(s) they speak, but has little or no effect on intelligibility.	Can articulate a high proportion of the sounds in the target language clearly in extended stretches of production; is intelligible throughout, despite a few systematic mispronunciations.  Can generalise from their repertoire to predict the phonological features of most unfamiliar words (e.g. word stress) with reasonable accuracy (e.g. while reading).	Can employ prosodic features (e.g. stress, intonation, rhythm) to support the message they intend to convey, though with some influence from the other languages they speak.
B1	Pronunciation is generally intelligible; intonation and stress at both utterance and word levels do not prevent understanding of the message. Accent is usually influenced by the other language(s) they speak.	Is generally intelligible throughout, despite regular mispronunciation of individual sounds and words they are less familiar with.	Can convey their message in an intelligible way in spite of a strong influence on stress, intonation and/or rhythm from the other language(s) they speak.

Table 7.2c. Common European Framework of Reference for Languages - *Companion Volume* “phonological control descriptors” (Council of Europe, 2001: 28)

In my opinion, the most important modification to the new phonological control descriptors is the separation of segmental and suprasegmental features, which are referred to as “sound articulation” and “prosodic features”. Word and sentence stress, intonation, and rhythm are all mentioned, and the words “intelligible” and “intelligibility” appear several times. Although these descriptors are still not language specific, they are, I believe, a vast improvement on the previous descriptors, and they remain usable by teachers and learners alike. “Usability”, or “ease of use” (Harding, 2016: 13) is the factor which must be balanced with level of detail: the more elements that a set of descriptors mentions, the more pedagogical value it has, the fewer elements mentioned, the easier it is to use. The pilot studies which I conducted in 2015-2016 (unpublished) with trainee teachers and university colleagues who used the prosody descriptors with their students showed that they were usable, but only after training, and highlighted their usefulness in better understanding pedagogical objectives.

### 7.3. Towards an embodied and metacognitive approach centered on prosody

Having examined some of the issues surrounding goal setting, we will turn to questions concerning the methodology of teaching English pronunciation focusing on prosody. I started in the introduction by framing this report in terms of embodied cognition, and in this section, I would like to look at the pedagogical implications of the theory we saw in part one. I will begin by presenting a few other examples of embodied learning techniques from the 1950s to the present day, then I will present some of the research on music and language learning, and finally we will briefly discuss the importance of metacognition and strategy development in language learning and teaching.

#### 7.3.1. *Examples of embodied approaches to language learning*

In this section, I would like to briefly present a number of pedagogical approaches which fall under the umbrella term of “embodied learning”, in that they rely heavily on the movement of the body or parts of the body, and although they can be done sitting down, they usually involve standing up. The Centre for Educational Research and Innovation defines embodied learning as follows: “(e)mbodied learning refers to pedagogical approaches that focus on the non-mental factors involved in learning, and that signal the importance of the body and feelings” (OECD, 2018: 117). Embodied learning has two requirements, the use of bodily activity, and the integration of tasks, in other words: whether bodily activities are related to learning tasks in a meaningful way (Skulmowski & Rey, 2018: 1). Some of the approaches I will present here are considered as complete methodologies for teaching language, some

are to be integrated into teaching practices. They do not all focus specifically on pronunciation in and of themselves, and I do not personally advocate strict adherence to any *single* methodology. All of the approaches I mention have been shown to improve learning outcomes in studies, but no one approach can be considered a panacea for pronunciation instruction, as all learners and learning situations are different. For this reason, I would advocate drawing on the aspects of any of these approaches that a teacher deems appropriate for their learners in a given context. The principle I do advocate is to engage the body in the process of learning pronunciation: engaging the body to better understand and reinforce aspects of prosody and phonemes, and to integrate the more physical activities into an approach which includes metacognition activities such as guided discussion, and more traditional learning activities.

One of the most well-known embodied approaches to language teaching is Total Physical Response (Asher, 1969). It is usually practised with beginner learners, and is particularly recommended for children, although in Asher's original study (using American learners of Russian), the adult group actually outperformed the children's group (Asher, 1969: 16). In this approach, the emphasis is on listening, but with a view to developing oral production skills later on. The teacher plans the lessons around vocabulary, and to a lesser degree, grammar. The lesson then consists in the teacher giving commands to the learners, who respond by carrying out actions (Richards & Rodgers 2001: 75-76). This approach is particularly interesting, because although it does not specifically target pronunciation, it places reception skills before production skills, and in this regard, it may be used to raise awareness to certain phonological features, especially with younger learners and / or lower levels.

The verbotonal system (Guberina, 2013), which I presented briefly in section 5.6 was developed in the 1950s and 1960s by Petar Guberina, and is quite widely used in the community of LX French teaching (Borrel & Salsignac, 2002; Sauvage & Billières, 2019; Bourreux et al., 2021) where it is usually referred to as *la méthode verbo-tonale (MTV)*. The verbotonal system applied to learning languages focuses on prosody, and works on retraining the ear, and on pronunciation by engaging the whole body. The verbotonal system is also used in the teacher training MOOC for LX French teachers produced by *France Université Numérique* (Billières et al., 2018).

The Silent Way (see also section 5.3) is a method developed by Caleb Gattegno (Gattegno, 1963), and widely used for teaching languages (Young and Messum, 2012), including pronunciation classes in English for specific and academic purposes in French universities (Herry-Bénil, 2011). It is a form of multimodal teaching, as the teacher points with a stick to a chart of coloured rectangles to represent whatever is being taught, and talks as little as possible, hence The *Silent Way*. In the context of pronunciation instruction, The Silent Way can be considered as an embodied approach, as the learners



are generally standing in a semi-circle around the board & teacher, not reading or taking notes, not interacting in any way, except looking, listening and responding physically to stimuli. The Silent Way is mainly used for teaching phonemes, but the coloured rectangles can be grouped together for example to represent syllables and tone units in order to work more on prosody.

Haptic pronunciation teaching (HaPT) is an approach developed by William Acton, and uses touch and gesture to present, correct, practise or anchor new or repaired sound structures (Acton et al., 2012: 235). The approach has been developed over decades of work with Asian and other learners of LX English, beginning with his work in the 1980s on “kinesthetic correcton” (Acton, 1984).

In the French context, other teachers have used gesture to represent sounds in a similar way to Acton, including Stéphane Soulaine, who used gestures to represent phonemes in his PhD, creating what he called a *gestographie* (Soulaine, 2014). Soulaine continues to use an embodied approach to English teaching at Montpellier University on the Master’s programme to train LX French teachers and researchers, using techniques from dance choreography and theatre (Soulaine 2018). Choreography is also a space where theatre and dance overlap, and in the context of teaching English in France, several teachers have used embodied techniques drawing on dance and theatre, such as Jean-Rémi Lapaire, who uses the choreography of dance and theatre to help both adult and younger learners of LX English. Lapaire has used total body movement inspired by dance and theatre to teach English grammar (Lapaire, 2006; 2014b) and also to explore lexical and grammatical expression both in LX English and French (Lapaire, 2010). Finally, theatre in particular is widely used in France for teaching LX French, English and other languages, particularly in secondary schools (Schmidt, 2006, Aden & Eschenauer, 2020; Eschenauer, 2021; Eschenauer et al., 2022; Bourrain et al., 2021), but also in primary schools (Schmidt, 2010) and in universities (Privas-Bréauté, 2018; Tummillo, 2018). There are also several organisations teaching English through theatre in France, mainly in schools, such as *Langues en Scène* and *Drama Ties*.

### 7.3.2. *Music, prosody and language learning*

As we have already seen (chapter one), music and human speech share many features, prosody in particular being very close to music in the way it structures language. Musical annotation has often been used to explain prosody, even as far back as the 18<sup>th</sup> century, with Joshua Steele’s use of musical annotation (Steele, 1779), indeed the term prosody itself comes from the way the Ancient Greeks taught song. Music is often used in the language classroom, usually for grammar or vocabulary reinforcement, or for listening comprehension, however, I am particularly interested here in the effects of music and song on language learning. It has been shown that musical priming (in the shape



of perception tasks) can have a cross-linguistic effect on language production (Jungers et al., 2016), which will be no surprise to teachers who have used music in the language classroom or examined the effect of its use in research. The benefits of music in learning languages, particularly regarding prosody, have been demonstrated by numerous studies, for example on the perceptual level, in a study involving stress processing in English by L1 French speakers (Kolinsky et al., 2009), it was found that the musically trained group outperformed the non-musically trained group significantly.

Musical ability has been shown to be similar to auditory perception (Caroll, 1993), and clearly has both sensory and cognitive components (Visser et al., 2006). Many studies have found positive correlations between musical aptitude and other cognitive and social skills, including linguistic skills (Milovanov et al., 2008: 81). Singing talent and the ability to imitate language sounds have been correlated in several studies, most notably Robert Slevic and Akira Miyake's (2006) study, which was the first serious study to confirm the musical-ability hypothesis. Using 50 Japanese learners of English in America, their study correlated musical ability with proficiency in LX phonology (Slevic & Miyake, 2006: 679).

Christiner and Reiterer (2013) also positively correlated several abilities, including singing talent with comprehensibility in a language imitation task, considering prosodic features such as stress and intonation. And not only when it comes to processing the linguistic information conveyed by prosody in a foreign language. It would seem that the training effect is greater using singing, when identifying the emotional information contained in prosody, both in familiar and unfamiliar languages; in a study comparing the ability of 6-7 year-old children to identify emotions in familiar and unfamiliar languages, the children trained in singing could distinguish happy/sad and angry/fearful in tone syllables (created digitally and void of any segmental features) better than those children trained using drama or keyboards only. Regarding production, in one study (Ludke et al., 2014), participants who sang considerably outperformed the groups who were speaking or speaking considerably when learning a new language. As part of her PhD study, which I co-supervised during the last year, Sayena Molaie (2018) conducted a study using engineering students. She identified the most musical-rhythmic students, and then developed tailored activities to the "musical-rhythmic" group. In a pre-test / post-test protocol she compared their results with a control group, and her results showed that for the students in this context, activities tailored towards musical-rhythmic students had a positive outcome both in learning vocabulary, and in motivation to learn. It is also a feature of music that it can change our mood, for example, Fonseca-Mora and Machancoses (2016) detail the benefits on emotional well-being of music, and the link between a sense of well-being and effective learning. Patel (2011) suggests in his OPERA hypothesis that musical training benefits not only auditory perception, but also reading. His hypothesis contains five conditions:

- 1) **Overlap**: there is anatomical overlap in the brain networks that process an acoustic feature used in both music and speech (e.g., waveform periodicity, amplitude envelope),
- (2) **Precision**: music places higher demands on these shared networks than does speech, in terms of the precision of processing,
- (3) **Emotion**: the musical activities that engage this network elicit strong positive emotion,
- (4) **Repetition**: the musical activities that engage this network are frequently repeated, and
- (5) **Attention**: the musical activities that engage this network are associated with focused attention. (Patel, 2011: 1)

According to the OPERA hypothesis, “when these conditions are met neural plasticity drives the networks in question to function with higher precision than needed for ordinary speech communication” (Patel, 2011: 1).

In the second chapter, we saw how human beings were capable very early on in their development of perceiving beats, and later on in synchronising their movements to these beats. In dance and in music, the concept of beat perception and synchronisation is of course essential, as it is in language. In an earlier paper, Patel (2006) explains the “beat perception and synchronization hypothesis”, which states that vocal production learning (VPL) is a prerequisite for humans, and other species, to be able to extract a *pulse* from periodic acoustic events (like an internal metronome), and use this inferred pulse to synchronize movements to these external events in a predictive and flexible way (rhythmical entrainment). In the context of France, music has often been used to teach LX French, with approaches which put prosody first and foremost, such as Grégory Miras’ PhD study (Miras, 2014), or Régine Llorca’s “*ritmimots*”, a similar approach to Carolyn Graham’s jazz Chants (Graham, 2000). The five conditions Patel mentions in his OPERA hypothesis, and the principal of rhythmical entrainment are very much at the heart of many of the activities which I have been using for many years now when teaching pronunciation, especially since my involvement in the ministry of education-funded *Innovalangues* project. During this project from 2012-2016, I was part of a group which I later headed, and which was composed of teachers, mainly of English, but also of Italian and French. We all had “amateur” interests in performing arts which fed into our teaching, including music, singing, poetry, dance and theatre. The group in question was tasked with producing resources and activities to help learners improve their oral production, hence the group’s name, *THEMPPO* (*THÉMatique Prosodie – Production Orale*). We were able to develop and experiment with various activities and digital tools, all based around the idea that prosody was at the heart of oral communication, and that to work on prosody, the body and mind had to be engaged together. The approach we developed borrows heavily from various activities in those performing arts, and combines them with guided metacognition to help learners develop strategies (Frost & Picavet, 2014), as we shall see below. In an article I co-authored in French with one of the members of the group, who is an English teacher, a musician, a performing

poet, and has a PhD in musicology (Frost & Guy, 2016), we referred to our approach as “*musico-prosodique*”, however this underplays the importance of metacognition which the activities help to develop.

### 7.3.3. *Metacognition and learning*

Broadly speaking, metacognition may be defined as “thinking about thinking”, or “thinking about learning” in the case of education. Metacognition has been linked strongly by many studies to positive learning outcomes (Wang et al., 1990). Metacognition in language learning helps learners to develop strategies to better understand and organise language input in various forms to aid the learning process, such as planning, solving problems, and repairing errors (Proust, 2021). Rebecca Oxford (1990: 8) defines learning strategies as “specific action taken by the learner to make learning easier, faster, more enjoyable, more self-directed, more effective and more transferable to a new situation”. She classifies strategies into direct strategies, and indirect strategies. Direct strategies directly involve the target language, and these include memory, cognitive, and compensation strategies, all of which require mental processing of the language (Oxford, 1990: 37). Indirect strategies include three sets of strategies, including metacognitive, social and affective strategies (Oxford, 1990: 151). There may be, of course, a great deal of crossover between these categories of strategies, and the cognitive and affective dimensions of learning are also hard, or even impossible to separate, as Vygotsky (1978) argued. Marnie Reed and Christina Michaud (2015: 454) refer to metacognition as “the missing link between theory and practice”, as too often, pronunciation instruction involving suprasegmental features such as intonation remains at the level of production, and not metacognition.

Graeme Couper (2006; 2009; 2022a; 2022b) promotes the use of the development of metacognition through targeted discussion with learners around certain language points, particularly the stress and rhythm of LX English. Couper uses guided discussion to let terms emerge from the learners in what he calls Socially Constructed Metalanguage (Couper, 2022b: 2). Helen Fraser also uses metacognition by encouraging discussion about teaching and learning pronunciation in the context of teacher training (Fraser, 2006). Both Fraser and Couper’s work involves a similar process, whereby the teacher first exposes learners to complex prosodic and other phonological features, and through a gradual process of discussion, help each other to understand learning difficulties and find appropriate solutions. This process is close to the Vygotskian concept of languaging, defined by Neomy Storch (2017: 73) as the “[v]erbalisation of thinking processes in problem-solving activities”. Languaging can be self-directed or directed towards others, and as part of a guided discussion process, it can be integrated into various stages in the learning process. In the metacompetence-oriented model of phonological acquisition

proposed by Magdalena Wrembel (2007), she integrates metacompetences into the “facilitating device”, as we can see in figure 7.4.3:

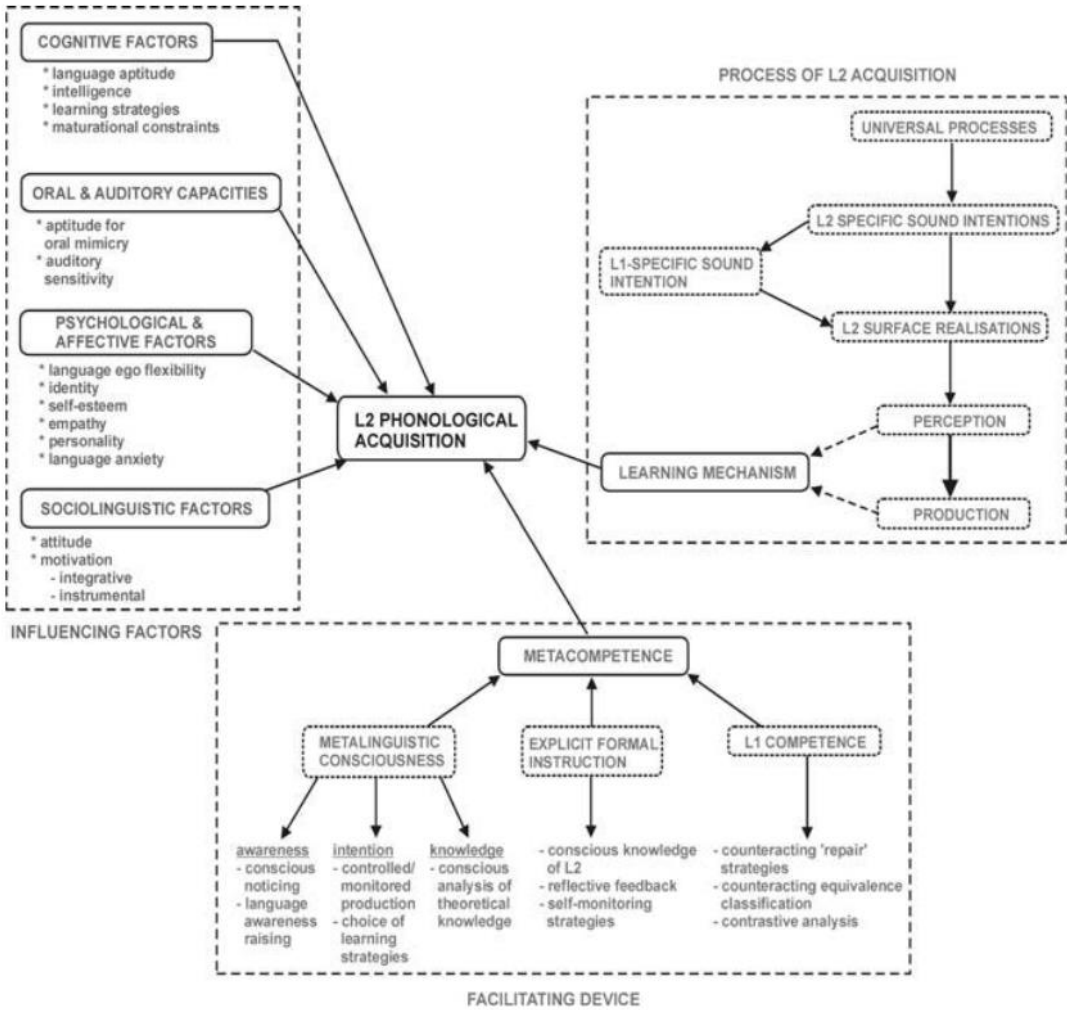


Figure 7.3.3. Metacompetence-oriented model of phonological acquisition (Wrembel, 2007: 191)

The three components of metacompetence cover many of the subjects which guided discussion with a teacher in the role of a mediator would help to develop. This form of dialogic scaffolding, a cornerstone of Vygotsky’s (1978) socio-cultural approach to teaching and learning. This process of exchange or “collaboration with more capable peers” (Vygotsky, 1978: 86) would focus on the difficulties learners have with producing intelligible speech, with comprehension of spoken English, through exploring the nature and role of prosodic features in English and in French (or other L1s).

7.4. An embodied and metacognitive approach centered on prosody: what and how

As I have mentioned several times during this report, for example when discussing isochrony (section 1.1.4), and the differences in focus and intonation between English and French (section 4.1.5), prosody

is as much a set of perceptual features as it is a set of acoustically measurable features. Prosody is an acoustic reality, but it is the effect it produces on the listener's ear which counts for a learner's spoken English to be intelligible so that communication can take place effectively. As my research developing the prosody descriptors which I presented in the previous section has shown, French learners, even at lower levels, usually know *where* to place lexical stress in English, but they usually do not manage to muster the correct balance of all the acoustic cues in order to produce the effect of correct stress on the listener's ear, even at more advanced levels of C1 and C2. And tonic stress and intonation are even more problematical for French learners, in that they often realise a rising or falling pattern, but not in the same way an L1 English speaker would. For these reasons, I suggest that in order to address the question of prosody in English with learners, particularly learners whose L1 prosody is as different from English as is the prosody of French, an approach is necessary which is both physical and cognitive. French learners do not only need to *know* which syllables should receive stress and which intonation pattern to use, but how to physically realise these prosodic forms. They need to *feel* them. And this is true on the level of production and on the level of perception. This is why I have adopted an embodied cognition approach which foregrounds prosody, engages the body and uses metacognition, as I will describe in this section.

In keeping with principles of an embodied cognition approach to learning and teaching LX pronunciation, we will consider the learners' minds, bodies and environments as forming a whole complex and interactive system: the physical environment of the learning situation and learners' own physical activity will have an effect on the way they think, feel, and learn. Physical activities will involve learners physically moving parts of their bodies, or indeed their whole bodies, and also drawing on the senses of sight, touch and hearing in as many different ways as possible, in a truly multimodal approach to pronunciation instruction.

#### *7.4.1. Towards a prosody-centered syllabus*

The actual content of an instructional programme will depend on the constraints of each learning situation. The principles and many of the points laid out in this section would apply to most learners of English, but would of course vary according to factors such as learners' level, L1, needs and objectives, the skills and knowledge of the teacher, and other environmental factors. As I mentioned in the introduction, and as we saw in the previous chapter, the question of whether to focus on segmental *or* suprasegmental features is a sterile one: both contribute towards intelligibility (Wang, 2022: 199), and both should therefore be taught. Prosody and phonemes should be taught as part of a complex, integrated, and interactive system (Derwing, et al., 1998; Zielinsky, 2015; McGregor & Reed,

2018). I do not therefore recommend teaching *only* prosodic features, but giving them more importance, and highlighting their effect on phonemes.

The following recommendations follow a contrastive phonological analysis of French and English, and are in accordance with the principles of optimality theory, markedness differential hypothesis, functional load, and a detailed analysis of the literature regarding learnability of phonological features and their impact on intelligibility, perception and comprehension. I suggest that foregrounding prosody in the content of a programme is primordial, especially if the learners' L1 is French, or any language where stress plays a less discriminating role, as "accentuation and intonation have a controlling function over syllables and segments". (Missaglia, 1999: 551), as we saw in section 6.3.2. Similar approaches have been put forward by various authors, such as Kenworthy (1987), Pennington's (1989) "top-down" approach, or Firth's (1992) "zoom" principal, used by Derwing, Munro and Wiebe, (1998). In the context of LX English and French learners, other authors have highlighted the importance and usefulness of focusing on prosodic features (Cooke, 1993; Arleo, 1995; 2013; Bertin, 2000, Herry and Hirst, 2002; Stenton & Gimeno, 2011). The syllabus or programme I suggest for French learners of English is essentially the components of the prosody descriptors (section 7.2):

1. Rhythm & stress: the nature and role of lexical stress & focus.
2. Reduced syllables: especially schwa but also /ɪ ʊ/ and, according to time, level, etc. the final /i/ and the syllabic consonants /ŋ n m/).
3. Stressed and unreduced vowels: all the English vowels, especially emphasising the differences between the nature and distribution of lax vowels /æ e ɪ ɒ ɑ: ʌ ʊ/ and tensing vowels /eɪ i: aɪ əʊ (j)u:/.
4. Connected and running speech phenomena: thereby integrating consonants with prosodic phenomenon, rather than teaching them apart, such as deletions, assimilations, resyllabification, palatalization.
5. Intonation: it can be introduced with phrasing and focus, or sentence stress early on, then returned to regularly, with emphasis on the form and meaning of different intonation patterns.
6. Allophones: as they occur in different varieties of English (such as SSBE, GA, relaxed forms of L1 English and world Englishes should be taught and practised as appropriate, for example the allophones [t<sup>h</sup> r ?] for /t/, [ɹ w] for /l/, [f t s] for /θ/, and [v d z] for /ð/.<sup>15</sup>

The teaching of segments would therefore be covered, but through the influence of prosody on phonemes: unstressed vowels, then stressed vowels, and consonants through the prism of connected

---

<sup>15</sup> See Murphy, (2014: 267) appendix A for a list of "possible candidates of intelligible and comprehensible non-native English speakers accessible on the Internet".

speech phenomena, and the presentation and practice of allophones in the context of varieties of English. The points in this programme may be covered superficially in a few hours with higher level learners, but much more time is required to discuss these features following activities, and to integrate them into activities involving spontaneous speech.

#### *7.4.2. Activities: principles*

In this section, I will briefly list some pedagogical principles based mainly on the research which I have presented in the previous chapters. There is very little in education which is truly new or original, and unfortunately, pedagogical trends are not always based on empirical, scientific findings. The principles which I lay out here are the result of the research which I have synthesised in this report.

The process of metacognition begins at the very start of a new course: the starting point for any new class is for a teacher and learners to get to know each other. It is essential for a teacher to understand the knowledge and skills of their learners, and to gain insights into their wider environments in order to plan content and activities. Group discussion activities and guided discussion around the experience of using and learning language is not only as a needs analysis, but also enables many of the foundations of the content of a pronunciation programme to be laid. This initial stage enables teachers and learners to negotiate realisable goals and learning objectives together, within the constraints of institutional requirements. This initial process of getting to know learners, also aids the development of appropriate strategies for continuing effective autonomous learning outside the classroom and after the end of a course.

The most important principles are to integrate the teaching of phonemes into teaching prosodic features, and in keeping with embodied cognition, to engage the body wherever possible in a meaningful way. Before working on pronunciation, if time allows, a warm-up with breathing exercises can aid relaxation, raise awareness to articulatory processes, create an atmosphere which is propitious to pronunciation instruction and improve learning outcomes (Carrera, 2021). It is important to work on reception *and* production together (Cauldwell, 2013): usually perception work before production work is most profitable (Patel, 2011; Jung et al, 2017), but production and guided discussion before listening tasks has been shown to produce positive results (Roussel & Tricot, 2015). Multimodality and developing multimodal literacy (Crawford Camiciottoli & Campoy-Cubillo, 2018; Dressen-Hammouda & Wigham, 2022) should be a priority, and this involves working on lower-order cognitive tasks related to perception and segmentation in addition to higher order cognitive tasks linked to linguistic and pragmatic processing. And whenever possible, activities should be followed by short guided discussions to help to develop strategies to improve learning.

There are many general pedagogical principles which have been proven to work in research in pronunciation instruction or in other aspects of language learning, and which are applicable to a prosody-centered and embodied cognition approach to pronunciation instruction. Pronunciation instruction should include inductive teaching (Kaulfers, 1929), intelligible and comprehensible input and more challenging input, progression from more controlled activities to free production activities within lessons and over a course, and a spiral approach to programme content (Jerome, 1960).

Finally, an essential principle of any teaching is that at least some of the activities are fun. If learners enjoy themselves, they are more likely to engage with the process of learning, and high motivation, strong self-efficacy, and skilful self-regulation all go together to help benefit as much as possible from pronunciation instruction (Moyer, 2014).

#### 7.4.3. *Activities: examples*

It is not my intention here to reproduce an exhaustive list of activities, nor to explain how to put the activities into practice, my aim is simply to illustrate some of the principles listed in the previous section with some concrete examples. Many of the activities fit into a PPP (presentation, practice, production) approach to progression within a lesson (Harmer, 2007: 66), or an ESA (engage, study, activate) lesson plan (Harmer, 2007: 67). Most of the activities engage the body in some way, and engaging the body can be a way to raise awareness, to practice, and to reinforce learning. The activities in this section are usually followed by some degree of **guided discussion** around the issues addressed by the activity in question, with the aim of helping learners to develop appropriate strategies.

**Warm-ups** are an essential part of an embodied approach to pronunciation teaching. It may only be a minute or two, and it may not even involve standing up, but typically it will involve adopting a correct **posture**, standing (or sitting) with the back straight and head up, thus enabling the vocal apparatus to work more efficiently. **Breathing** for a few cycles together focuses the mind and relaxes learners, all of which can aid learning. To warm up the voice, there are many **vocal warm ups** which are popular in theatre, and which can be used before working on pronunciation (Anderson, 1977; Berry, 1991; Maley, 2000; Shewell, 2009; Fisher & Kayes, 2016). It is particularly important to relax the muscles of the jaw, face and tongue and encourage learners to **adopt articulatory settings for English** (Honikman, 1964) which are appropriate for English, rather than French, such as more lip protrusion with rounding, more alveolar ridge and less upper teeth contact for /t d/, etc. Warm-up activities before lessons concentrating on rhythm can involve **rhythmical synchronisation** activities, such as walking in step (if space allows), and any number of **handclapping games** (Arleo, 2013). Other forms of **alignment activities** include **mirroring gestures in pairs** – this sort of activity is useful as a warm-up before



working on dialogues and role play in pairwork activities, for example when practising focus and intonation.

When working on **stress**, initially it is useful to **build awareness of the four acoustic correlates to stress** and their percepts. Using a battery of exemplar words which represent common stress patterns and reductions, which are frequently used by the learners in question, which are often mispronounced (*uniVERsity; demonSTRAtion; deVELopment, CHOcolate, etc.*): with individual and choral repetition **exaggerate each of the cues** in turn in choral and individual repetition, trying to make syllables longer without making them louder or higher, or higher, but not longer or louder. In guided discussion after an activity, it soon becomes apparent that all the cues work together. **Backward building** or **forward building** is also a useful technique for pronouncing longer words with suffixes and prefixes. These activities can be an **introduction to the discovery of stress patterns** (strong endings, etc.), and of the **laxing and tensing patterns** for stressed vowels. They are also a chance to **practise reducing syllables** to schwa and other vowels or syllabic consonants. All of these activities may be accompanied with **clicking fingers, clapping hands, tapping toes or tables** to reinforce the rhythms of English, and in discussion afterwards, the differences between the stress of English and the learners' L1 or L1s can be highlighted.

For work on **intonation and focus**, learners can practise **coordinating body movements** (eyebrows, fists, fingers, head-nodding, etc.) with the large intonational shifts which indicate focus, or sentence stress. This may be achieved in a variety of ways, but it is essential that learners *feel* the movements as they produce the language themselves, with “**mm-mm-mm**” “**tum tee tum**”, or “**la-la-la**” noises, and with actual language, with much repetition (individually, in pairs and small groups, and all together). The teacher can serve as a model, as can videos, particularly certain politicians, or people who have been trained in public speaking.

For **rhythm**, as we saw in section 1.1.4, English has often been described as “stress-timed” (Pike, 1945), and is marked by an evident alternation of strong and weak syllables. There is a continuum of discourse genres and types which can illustrate this, from prosodically rich discourse, to less prosodically regular spontaneous discourse, although even running speech also contains some very rhythmically regular sections. When working on prosody, begin with **prosodically rich and culturally important examples of prototypical speech** (Cauldwell, 1996; Lin, 2012), such as the examples given in section 2.3.1, for example **children's poems and nursery rhymes** and certain poetic forms, such as **limericks**, or **idioms, proverbs, insults, formulaic jokes** and other types of formulaic speech. Frequent training backwards and forwards along this spectrum, with the use of **gesture, finger and foot tapping, metronomes**, etc. all helps to reinforce the strong / weak alternation which characterises English.

When working on **phonemes**, it is important to **distinguish between stressed and unstressed vowels**, and within the stressed vowels, between **lax and tensing vowels**. The general principle is to **hyper-articulate at first**, and then make the movements more natural as sounds are integrated into words, words into longer utterances, and finally spontaneous speech. Activities to practice articulation of vowels include **silent speech guessing activities**, for example in pairs, learners mouth the names of fruits, vegetables, names, numbers, or any closed set of items while **hyper-articulating**. As learners become used to the format of the activity, they can try articulating more normally. This sort of activity not only helps develop articulation, but also helps learners to use visual cues, an important part of speech perception, and an important part of the detection of prosodically marked segments (Swerts & Kraemer, 2008; Dohen & Loevenbruck, 2009; Rapin & Ménard, 2019)

Finally, most of the activities I have mentioned so far involve speaking, listening and moving, but many of the problems which learners have with English comes from the fact that users of many other languages share the same alphabet as English, but have learned from a very early age to pronounce the letters in their own language, not with English sounds and rhythms (Stenton, 2009: 297). One of the ways to counter this, and to prepare less controlled semi-spontaneous speaking activities such as **presentations, role plays, theatre projects** – which can easily be filmed– is **prosodic annotation**. At its simplest, prosodic annotation may be simply **underlining the stressed syllable** when learners make a note of a new vocabulary item (which should be done as a matter of course), and perhaps **noting which syllables reduce to schwa**, or **marking the vowels or consonants which are apt to disappear** if a word is spoken quickly. Longer stretches of speech can also be marked for pauses and the tonal movements which accompany the focus can also be marked. There are many different ways of annotating prosody, and they all have their advantages and disadvantages, and it is a useful activity, once learners have seen a few, to discuss the merits of different forms of prosodic annotation and agree on a system or systems to use in class.

## 7.5. CAPT: embodied cognition and computer-assisted pronunciation instruction

I discussed different issues in computer-assisted pronunciation training (CAPT) in section 5.7, and it is not within the scope of this project to list the many digital tools which are available for learners. I have experimented with different forms of prosody visualisation in previous projects (Frost 2002; 2008), and use various digital tools in my teaching on a daily basis. This is a field which is changing so quickly, especially as smart technologies, artificial intelligence, big data, speech recognition and speech synthesis combine (Pennington & Rogerson-Revell, 2019: 370), and we have had brain to computer

interfaces that can recognise phonemes that users think without actually speaking for ten years now (Mugler et al., 2014).

In the context of an embodied approach to learning pronunciation, the onus is on the learner to move, and most of the digital technologies available for CAPT do not encourage this. In a teacher-guided, blended learning scenario however, there are several digital resources which may be used by learners to engage their body while learning about prosodic features. Learners can easily record video images of themselves with their smartphones or webcams, and view themselves critically. This can be combined with an active comparative approach, and imitation of videos from various platforms on the internet. Streaming services such as Netflix, provide subtitles, platforms such as TED Talks provide scripts, Youtube provides real-time captions, and Youglish enables learners to search for particular words and phrases: all of these tools can provide models for learners to imitate, and the use of video enables learners to compare visual cues such as the articulators and gestures. When working with particularly rhythmically regular speech, such as poems and nursery rhymes, an online metronome can help learners to adopt the strong/weak alternation of syllables in English, and learners can record themselves, listen critically, and then try the same procedure without the metronome (Picavet et al., 2013; Frost & Guy, 2016) clicking fingers or tapping toes with the beats. In this sort of “video active comparative” activity, learners can use technology to enhance the learning process in an autonomous setting, after class or after the end of a course, while engaging the body to support cognitive processes such as “noticing” (Schmidt, 2010).

Finally, videos can also be used for raising awareness, formal instruction, guiding learners to discover and practice language points and setting up activities. In 2020, I recorded 30 short pedagogical videos (five minutes long on average) as part of a funded (*IDEX*<sup>16</sup>) project with several colleagues at Grenoble Alpes University. The videos are integrated into a 24-lesson online learning path for students at the university, and are accompanied by interactive activities. The subtitled videos were recorded, edited and subtitled by a professional video engineer, and together with the activities, they cover all of the points in the previous section, and the learners are encouraged to engage in a number of physical activities, such as warm-ups, rhythm activities and video active comparative activities (Frost, 2022).

The use of digital technology, may provide positive learning outcomes, and constitute a useful complement to face-to-face teaching. Lee, Jang and Plonsky (2015) analysed 86 studies investigating the effectiveness of CAPT, and concluded that the effect was smaller than human-based pronunciation instruction. In addition, CAPT may be seen to increase the digital divide between those learners who

---

<sup>16</sup> *IDEX (initiative d'excellence)* is a system whereby teachers can receive funding from the French government via their universities if they are part of this programme.

can afford to use such technologies, and those who cannot, and the proliferation of online language learning services raises many other ethical questions.

## 7.6. The ethical dimension of pronunciation instruction

One of the most important questions for any teacher is that of our moral responsibility, and as language teachers, we must remember that the voice is an expression of identity, both individual and cultural. On the individual level, pronunciation work may make learners feel uncomfortable: discussing and observing the articulators such as the lips and the tongue is something which needs to be approached with sensitivity and caution. On a cultural level, English is the most widely spoken language in the world, and very often it is a *de facto lingua franca*. As teachers of English, perhaps we are contributing to its linguistic and indeed its cultural hegemony, and we should try to preserve and to learn the world's regional languages to offset this potential effect, as some authors suggest (Nathan, 2008). However, in the context in which most of my work takes place, English is very much a tool to complement rather than replace Learners' L1. Ethics in language teaching is still a relatively new field, and most of the literature deals with Teaching English as a Second Language (TESOL) not Teaching English as a Foreign Language (TEFL) (Thomas, 2009). However, as I mentioned in the introduction, those of us working in France are of course concerned with TEFL, not TESOL, so the ethical questions are often very different. I believe that by teaching English, we are giving learners something which will aid their social and geographical mobility, and in that respect, and speaking as a European, I believe very much in the principles of the Council of Europe, as outlined by their Language Policy Programme of the Education department in the *Common European Framework of Reference for Languages* (Council of Europe, 2001) and its *Companion Volume* (Council of Europe, 2020).

As I mentioned in chapter four, since 2000, most of the exchanges in English happen between non-L1 speakers, (Brumfit, 2001), so in Europe, as in most of the countries of the world, English is a tool for communication, and not a tool of colonisation, linguistic or otherwise. As a teacher in lifelong education, I often find myself beginning a course by taking this idea further when conducting a needs analysis of my new group of learners. Oral English is usually the area they ask to work on the most, and when we discuss why, it usually transpires that pronunciation – as a key factor in production, interaction and comprehension – is at the heart of what they need. Yet working on pronunciation can be ethically challenging. Our voice is, as we saw in chapter one, a strong social and psychological identifier. Encouraging someone to modify or to add to their accent is potentially modifying the way they express, project, and perceive themselves. Thomson and Foote (2019) set out the following ethical guidelines for teaching pronunciation:

1. Pronunciation instruction should primarily focus on intelligibility, rather than reduction of accent.
2. When teaching pronunciation, an L2 accent should be viewed as a natural part of L2 speech development; an L2 accent is not a speech disorder.
3. Individuals offering instruction should not make exaggerated claims about the efficacy of the instruction they offer, or the results of services or products offered.
4. Individuals or organizations offering pronunciation instruction should not use fear-based advertising that demonizes an L2 accent. Advertisements should be honest and appropriate.
5. Pronunciation instruction should not be continued if such instruction is unnecessary or ineffective.
6. Individuals offering pronunciation instruction should have specialized training in pronunciation; a degree in TESL or speech-language-pathology may not be sufficient to qualify someone as an expert of pronunciation.
7. Professionals offering pronunciation instruction should continue to seek professional development and be aware of new research developments in pronunciation research.
8. Professionals offering pronunciation instruction should respect the dignity and rights of all persons without prejudice as to race, religious beliefs, sex, gender identity/gender expression, socioeconomic status, sexual orientation, physical characteristics. (Thomson and Foote, 2019: 228-229)

This is a long overdue discussion, but let us not forget the importance of context. As a teacher trainer, I firmly believe that a well-trained and competent teacher is the best judge of the needs of their learners. The many assertions I have made in this report, the research I have carried out and the research I cite, are to be taken in the context in which I work – I would not presume to impose all of my conclusions on other teachers, researchers, and learners in other contexts. Regarding the first five points in Thomson and Foote’s guidelines, I believe I have justified my choices regarding the utility and efficacy of teaching pronunciation and focussing on prosody. As for points 6 and 7, I regularly accept invitations to train teachers, to speak in front of specialist and non-specialist audiences, and participate in the organisation of various forums to share good practice, here in Grenoble, elsewhere in France and abroad. The objectives of pronunciation teachers and researchers are changing, as are the goals of many educators – student-centered instruction is not a new concept, and nor are educational approaches which place the learner at the heart of a wider ecology of communities.

In the final section of this chapter, I would like to briefly present some of the research, teaching and teacher training projects I am involved with and some ideas for the coming years.

### 7.7. Research perspectives & projects

In this final section, I would like to outline the projects I am currently engaged in, and the directions I would like my future work to take. Firstly, however, I would like to briefly describe my research

activities in the university, and in the wider community of researchers in the field of applied linguistics in France and internationally.

I am an active member of several research groups (*associations* and *sociétés savantes*), and I am part of several scientific committees and reviewing boards for journals and annual conferences, so every year I participate in the organisation of at least two conferences, and review several articles or chapters. I also participate regularly in seminars in my research laboratory, Lidilem (*Laboratoire de recherche en Linguistique et Didactique des Langues Étrangères et Maternelles*), where I am part of two research groups: *EnForme* (*Formation des enseignants de/en langues*) and *DIDELOLE* (*Didactique et Linguistique de l'oral/oralité en langue étrangère*). The next *DIDELOLE* event, which I am co-organising, is a day of seminars and workshops on theme of embodied learning.

I currently co-supervise two PhD students: Hannah Burroughs, who is defending her thesis this year, and Stella Ville, who started her PhD this year. Hannah is comparing the effects of different approaches to language teaching on motivation and learning outcomes in France, Finland, and Canada. Stella is using the verbotonal system to investigate the effect of this embodied approach to teaching pronunciation with adult learners of English in Spain.

The research project PIC (Prosody, Intelligibility and Comprehensibility) is an ongoing project which brings together my main research and teaching interests, i.e. a prosody-centered approach to teaching and learning pronunciation, comparing face-to-face embodied practices and online autonomous learning, and measuring intelligibility. The learning path I created as part of a teaching project which I outlined in the previous section has now been completed by two successive cohorts of first-year students in applied foreign languages (*LEA*). This is a total of over four hundred students who have completed all twenty-four units and the pre-tests and post-tests. Alongside these students, a smaller group of about 30 students have agreed to complete the pre-tests and post-tests only, without watching the videos or completing the activities, and they constitute a control group, so I will be able to compare learning outcomes. The students have all agreed to participate in the ongoing study, and the recordings they make as part of their pre-tests and post-tests are part of a database of similar pre-test and post-test recordings from face-to-face teaching programmes. The recordings are comprised of readings of the same text, and with the help of a Master's student and some funding from the Lidilem, the recordings are being uploaded onto the online corpus management platform Ortolang, along with the necessary metadata to make the recordings useful and comparable. In order to assess the learning outcomes of the participants in this project, I carried out a pilot study (Frost, 2021) using the prosody descriptors (Frost & O'Donnell, 2018), and 10 expert raters. I hope in the future to collaborate with Sylvain Coulange, who developed a set of instruments using 16 acoustic parameters

to measure the rhythm of Japanese learners of French (Coulange & Rossato, 2020), and is currently adapting it for measuring English learners. This approach has already been tried with French learners of English (Herry and Hirst, 2002), but with limited success. As I explained in section 6.4, I believe that intelligibility, not comprehensibility, is the most relevant construct for pronunciation research, and as we saw in section 7.3, I believe that the most useful way of assessing intelligibility for establishing learning objectives is to measure the components which research has shown consistently contribute to intelligibility. By comparing a machine-based approach and a human-based approach to the assessment of prosodic features, I hope to improve our understanding of how prosody contributes to intelligibility, and that this will help improve our understanding of priorities for teaching English pronunciation in France.

Comparing the attainment of various learning objectives, for example by using controlled studies, by digitally manipulating speech, or by comparing human-based instruction or assessment, continues to provide us with useful quantitative data, which can help to establish an empirical basis for priorities in pronunciation teaching (McAndrews & Thomson, 2017). Quantitative studies generally entail some form of qualitative data collection, but it is generally subsumed by the quantitative aspects of the study. Quantitative data are necessary, of course, and I continue to adopt a mainly quantitative approach with regard to pronunciation assessment, using descriptors to help frame the necessarily subjective approach of human judgement. The principal reason that I use descriptors, and that I encourage trainee teachers to do the same, is to improve pedagogical practice; the Common European Framework, for example, is not primarily a tool for assessment, although that is what most learners and teachers think. The descriptors were established for pedagogical reasons, examples of which are to provide “a common basis for the elaboration of syllabuses”, “to overcome barriers to communication”, for stakeholders to “reflect on their current practice” (Council of Europe, 2001: 1). Teaching and learning pronunciation is an eminently human endeavour, and setting curriculum priorities is necessary, but as to *how* pronunciation is taught and learned, a more human-oriented approach, using qualitative methodologies, is necessary to compliment quantitative studies. I have administered many questionnaires and carried out many interviews over the past 25 years, trying to better understand the effects of various aspects of teaching on the cognitive and affective processes of learners, and this is often what has brought me the most insight into learning. There are currently longitudinal studies underway which explore the cognitive and affective aspects of learners within an embodied cognition framework in various contexts in France (Rouaud et al., 2022), and I would like to conduct more qualitative research, using a longitudinal approach, combined with a more case-study approach (Bower, 2017), with a smaller numbers of learners, providing rich, qualitative data to allow

me to explore certain questions more deeply, such as motivation, attitudes and thought processes around the issues of accent, intelligibility and pronunciation instruction.

Finally, in the last decade, an increasing amount of my work has involved teacher training: I have worked on the Master's programme for primary school teachers, I teach on the Master's degree for secondary school teachers, I teach classes for researchers for presenting work at international conferences, I teach language assessment for French teachers, I run a vocational diploma for English teachers, which I helped to create 10 years ago, and on which I teach several modules, I train primary and secondary teachers for content and language integrated learning (CLIL), and finally, I have also conducted dozens of teacher training workshops and seminars, and continue to do so regularly. Much of the work I do in teacher training involves pronunciation, and although research has shown us much of the procedural knowledge teachers need in phonetics and phonology, and the vocational knowledge they need in pedagogy, more research is necessary to understand *how* to put this knowledge together for teacher training programmes (Murphy, 2017). This research must explore teacher psychology (Mercer, 2018), but it must be relevant to the environment where the teaching happens. By adopting case study techniques and using longitudinal studies, such as Michael Burri and Amanda Baker's recent (2021) study at an Australian university, and teacher-focused reflective practice (Gilquin et al., 2022), I hope to improve our understanding of teacher's needs and how better to help them to achieve their aims in the area of teaching and learning pronunciation here in France and elsewhere.



## Conclusion (part two)

In this part we have focused on issues more directly concerned with language education, starting with an examination of the environment in which most of my teaching and research has taken place – France. While it is true that most of my learners are French, there are also many learners in France who do not have an L1 French background, which is why it is important to situate the French context within the wider context of World Englishes. We saw in chapter five that the French education system concentrates essentially on two models of English – Standard Southern British English and General American. Despite the recommendations of experts regarding the teaching of English in France, the actual time spent in teacher training colleges learning how to teach languages had reduced considerably in the last decade, and at the same time, English is being taught more often and to younger learners in primary education. On a more positive note, we have seen that the Common European Framework of Reference for Languages (Council of Europe, 2001; 2018) has had a positive influence on language pedagogy on the secondary sector. In the tertiary sector, the picture remains very mixed, and throughout the state education sector in France, digital resources are increasingly being used to make up the shortfall in human-based language teaching.

Chapter six presented research in the field of LX pronunciation teaching, particularly studies focusing on prosody. In nearly all of the studies, prosody was shown to correlate with intelligibility, and in studies measuring learning outcomes of pronunciation instruction, gains in prosody performance were consistently found. It is clear however, that concerning accent, intelligibility, and comprehensibility, these constructs are often poorly defined, and that the measurement tools employed by researchers are often not suited to the constructs being measured. Intelligibility and comprehensibility are often used synonymously, and differentiated by the way they are measured: Intelligibility is systematically measured by techniques such as orthographic transcriptions and comprehension questions (which rely on many linguistic features other than pronunciation), and comprehensibility is measured in the same way as perceived accentedness (with a Likert scale, often with naïve listeners). Several more recent studies acknowledge that comprehensibility relies on more linguistic processes than simply phonetic and phonological cues, but it is still a mainstay of pronunciation studies, and yet the result of comprehensibility tests is often a numerical value between one and nine, which is of limited pedagogical value. In chapter seven, I suggest that at the level of production, pronunciation research should focus on the construct of intelligibility, and that the most useful way to measure this construct, is by deconstructing it into the phonological and phonetic features which research has shown contribute to it. I would further suggest that the best way to do this is by using scales of descriptors. This is partly for practical reasons, but mainly because of the positive impact that using descriptors has on pedagogy, as it allows teachers and learners to identify difficulties and negotiate learning outcomes.

This is the principle of the Common European Framework's descriptors, but as they are not language-specific, and as each learning situation is different, I decided to develop a set of prosody-based descriptors (Frost & O'Donnell, 2018), which I am currently using in an ongoing project (Frost, 2022). Finally, the research I have presented in this project has led to the pedagogical principles and activities which I present in chapter seven. It is important to work on reception and production together wherever possible, as language perception and production are closely linked, as we saw in chapter two. At the level of perception, this includes exposing learners to as many different varieties of English(es) as their needs require, according to the affordances and constraints of the learning context. At the level of production, these principles include foregrounding prosody, engaging the body, considering the learners and their environment, and encouraging learner metacognition. In this way, learners can improve their overall pronunciation and add to their store of pronunciation features, which can help them with both understanding spoken English, and producing more intelligible English.

## General conclusion

### Motivation

As I explained at the outset of this report, I have been motivated by one question in my career since I arrived in France, namely how to help French learners (and other learners) better understand spoken English. My interest in researching and teaching pronunciation has been fuelled by this question, and this is why I have focused on the constructs of accent and intelligibility. I am of course interested in comprehensibility, and when I teach courses on general English, or English for specific and academic purposes, the learning objectives are not limited to pronunciation and intelligibility. I have specialised where I think my work can be the most useful here in France: by teaching and researching pronunciation in order to improve perception, production, and intelligibility. This means, of course, that I need satisfactory definitions of these terms. There are models of perception which I consider to be good and useful representations of the complex processes of perception (Liberman & Mattingly, 1985; Schwartz et al., 2012), of production (Levelt, 1989; 1999), and of LX use (MacWhinney, 2005; 2018). However, since I began to read and conduct research concerning the construct of intelligibility, I have often been unconvinced by the way in which a large body of research has defined this construct, and by the instruments with which it is often measured. For these reasons, I have proposed a model of how comprehensibility and intelligibility can be understood (chapter six), and how the factors which contribute to intelligibility can be assessed and learned (chapter seven).

### Context and content

If I have chosen to put prosody at the heart of my interests, it is because the context in which I work has dictated the content of my work. The research I have read and which I have conducted myself demonstrates that prosody is strongly linked to intelligibility. The studies I presented in chapter six of this report show that this is especially of European languages, particularly where Romance & Germanic languages meet. This is true of English & French, of course, but also phonologically similar pairings such as German and French, and Dutch or English and French. Even studies concerning Chinese and other Asian languages, where segmental phonetics and phonology are more closely correlated with perceptions of accent, find that duration consistently impacts intelligibility.

### Projects past and present

My own research has demonstrated that French and English listeners perceive stress using different cues (Frost 2011), that French learners realise prosodic features using different acoustic correlates to L1 English users (Frost & O'Donnell, 2018), and that the intelligibility of French learners' pronunciation can be assessed accurately and usefully (Frost 2021). I will always be interested in measuring learning

outcomes because I work in the field of education, and the creation of a learner corpus on a dedicated platform will help colleagues in France and elsewhere to do this. Much of my research in the past has involved learner metacognition – thoughts, attitude and motivation – and I would also like to understand more about teacher metacognition (Fraser, 2006, Couper, 2016; Burri & Baker 2021; Gilquin et al., 2022).

## Learning and teaching

Learner metacognition has long been known to have a positive effect on learning outcomes (Wang et al., 1990): it helps learners and teachers to better understand what learning objectives to fix and how best to achieve them. Learners themselves know more about who they are, and about how and when they use language than their teachers. Perhaps, in a world where learners have such ready access to material in all languages devices which accompany them everywhere they go, what we do in classrooms is less important than it was. As Purcell and Suter pointed out (1980: 84) “The variables which turn out to be important seem to be those which teachers have the least influence on”. We can however, as teachers, encourage our learners to focus their attention on the particular features which are most useful to their needs as individuals. Learner needs may involve more priority being given to intelligibility (involving a more accent addition approach), or maybe to accuracy (where high-stakes language use may imply explicit instruction towards one or more models), or more likely, a combination of both. In all of these scenarios, focusing on prosody has been shown to produce positive results, often in short periods of time, and even with older learners. Whichever learning objectives are set in a given learning situation, they need to be assessed – by learners themselves, by teachers, and by other stakeholders. The Common European Framework of Reference for Languages has brought many benefits in France and worldwide, and if the phonological control descriptors in the first version of the Framework (Council of Europe, 2001), or in the new *Companion Volume* (Council of Europe, 2018) are not a good fit for a given learning situation, then teachers, learners and researchers can make their own scales of descriptors.

## From embodied cognition, to embodied metacognition

Pronunciation - and especially prosody - is a cognitive skill, but it is inherently embodied. There are many means to achieve positive pronunciation learning outcomes, many ways to focus on prosody, and many methodologies and tools at the disposal of teachers and learners. Learners and teachers in any given context know best which to choose. I would suggest that the research presented here supports four principles for guiding stakeholders in making those choices:

1. Knowing the learners and their environment: this means understanding learners' individual and social needs, how they have learned languages, how they use languages, and how they want to use languages.
2. Working on reception *and* production wherever possible, and use as many varieties of English as possible.
3. Foregrounding prosody: segmental and suprasegmental features must both be learned, and in an integrated approach, prosody should be foregrounded, and when phonemes are targeted, the way they are influenced by prosody deserves consideration.
4. Engaging the body as much as possible: this means standing up, warming up, emphasising the importance of the visual cues to language (facial expression, gesture, and the visible movement of articulators).
5. Encouraging metacognition: this can be achieved by following embodied activities with guided discussion and other activities which actively encourage metacognition. This helps learners to better understand and set learning objectives, to develop appropriate strategies to improve learning outcomes.

### Teacher training

In the general introduction and in chapter six, we saw that there has been a renewed interest in pronunciation research, and that this is beginning to translate into changes in teaching practices, but the change is slow. Despite experts' recommendations and ambitious targets in language education policies in France, a lack of investment in teacher training at the national level has led to less teacher training in pronunciation over the last decade. At a local level, in my capacity as a teacher trainer, I intend to continue emphasising the importance of research-based practice when teaching pronunciation. In order for teachers to help their learners to teach English pronunciation effectively, in addition to skills in pronunciation pedagogy, a certain degree of formal knowledge of phonetics and phonology is required. This is particularly true for English as the target language, but knowledge of the learners' first language or languages is also necessary, so it is important to teach phonetics and phonology in modern language departments in French universities and teacher training colleges.

### Engagement, entrainment, and *engagement*

In embodied cognition, mind, body and environment are part of a system to be considered as a whole, and embodied teaching using metacognition provides the means to do this. Learners in France - and elsewhere - are often disengaged with their learning, however, it is difficult not to engage with an activity if you are standing up, moving around, and in the middle of a group of one's peers who are all doing the same. The research presented here has shown the relevance of social alignment, rhythmical

entrainment, and neural entrainment to learning language, and pedagogical activities inside and outside the classroom can harness these processes to aid learners.

In a wider sense, an embodied cognition approach not only helps learners engage with their learning, but it also provides a framework for researchers, teachers and learners to engage with other stakeholders in the wider environment. Institutionally, research in teaching foreign, second and other languages sometimes struggles to find recognition in French universities and research structures (Savatovsky, 2011). I hope this approach can also help to provide some *engagement* (in the French sense) at a local level, and with national and international partners.

Finally, every context is different, and I would not pretend any external validity for my research – unless the prosodic properties of the L1 speakers are also very different from English, however, I hope the research I have presented here is at least relevant to my learners and colleagues in France today.

## Bibliography

- Abel, C. (2019). La formation continue en didactique de la prononciation – un outil pour dépasser les querelles méthodologiques ? *Recherches en didactique des langues et des cultures*, 16(1). <https://doi.org/10.4000/rdlc.4439>
- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh University Press.
- Abou Haidar, L. (2021). L'oral à l'ère du numérique : enseigner et apprendre autrement ? *ALSIC (Apprentissage des Langues et Systèmes d'Information et de Communication)*, 24(2), 20. <https://doi.org/10.4000/alsic.5739>
- Abou Haidar, L, Lorca, R. (Eds.) (2016). L'oral par tous les sens : de la phonétique corrective à la didactique de la parole. *Le Français dans le monde. Recherches et application*, 60.
- Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly*, 18(1): 71-85.
- Acton, W. (2023). Haptic Pronunciation Teaching (HaPT). *Acton Haptic Pronunciation*. <https://www.actonhaptic.com/hapt>
- Acton, W., Baker, A., Burri, M., & Teaman, B. (2013). Preliminaries to haptic-integrated pronunciation instruction. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference, Aug. 2012* (pp. 234-244). Ames, IA: Iowa State University.
- Aden, J., & Eschenauer, S. (2020). Une pédagogie enactive-performative de la translanguageance en milieu plurilingue. In B. Schädlich (Ed.), *Perspektiven auf Mehrsprachigkeit im Fremdsprachenunterricht – Regards croisés sur le plurilinguisme et l'apprentissage des langues* (pp. 177–199). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-662-61770-0\\_8](https://doi.org/10.1007/978-3-662-61770-0_8)
- Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training. <https://doi.org/10.1007/s10639-019-09955-7>
- Akita, M. (2005). The effectiveness of a prosody-oriented approach in L2 perception and pronunciation training. *Gakujyutsu Kenkyuu*, 53, 1–22.
- Alazard, C. (2013). Rôle de la prosodie dans la fluence en lecture oralisée chez des apprenants de Français Langue Étrangère. [PhD. thesis], Université de Toulouse 2.
- Alazard, C., Astésano, C., & Billières, M. (2010). The Implicit Prosody Hypothesis applied to Foreign Language Learning: From oral abilities to reading skills. *Speech Prosody*, 1–4. [https://www.isca-speech.org/archive\\_v0/sp2010/papers/sp10\\_648.pdf](https://www.isca-speech.org/archive_v0/sp2010/papers/sp10_648.pdf)

- Alnafisah, M., Goodale, E., Rehman, I., Levis, J., & Kochem, T. (2022). The impact of functional load and cumulative errors on listeners' judgments of comprehensibility and accentedness. *System*, *110*, 1–14. <https://doi.org/10.1016/j.system.2022.102906>
- Allen, G. D. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, *3*(2), 75–86. [https://doi.org/10.1016/S0095-4470\(19\)31351-8](https://doi.org/10.1016/S0095-4470(19)31351-8)
- Altmann, H. (2006). The perception and production of second language stress: A cross-linguistic experimental study. [PhD. thesis], University of Delaware.
- Altmann, H. & Vogel, I. (2002). L2 acquisition of stress: The role of L1. (Paper presented at the DGfS Annual Meeting, Mannheim, 27 February-1 March 2002).
- Anderson, V. A. (1977). *Training the speaking voice*. Oxford University Press.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure. *Language Learning*, *42*(4), 529–555. <https://doi-org.sid2nomade-1.grenet.fr/10.1111/j.1467-1770.1992.tb01043.x>
- Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal*, *11*(4), 5–22.
- Archibald, J. (191) *Language Learnability and Phonology: The Acquisition of L2 Metrical Parameters*. [PhD. thesis], University of Toronto.
- Archibald, J. (1992). Transfer of L1 parameter settings: Some empirical evidence from Polish metrics. *Canadian Journal of Linguistics*, *37*, 301-339. <http://doi.org/10.1017/S0008413100019903>
- Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics* *35*(1). 167–181. <https://doi.org/10.1515/ling.1997.35.1.167>
- Arleo, A. (1995). It don't mean a thing if it ain't got that swing... Accentuation, rythme et langue de spécialité. *Recherche et pratiques pédagogiques en langues de spécialité - Cahiers de l'APLIUT*, *14*(3), 9–26. <https://doi.org/10.3406/apliu.1995.3461>
- Arleo, A. (1997). Counting-out and the Search for Universals. *The Journal of American Folklore*, *110*(438), 391–407. <https://doi.org/10.2307/541665>
- Arleo, A. (2006). Do children's rhymes reveal universal metrical patterns? In P. Hunt (Ed.), *Children's Literature: Critical Concepts in Literary and Cultural Studies*, vol. IV (pp. 39–56). Routledge.



- Arleo, A. (2013). Trying to Make It Real: Harnessing Foreign Language Teaching to Children's Folklore, Formulaic Language and Rhythm. *E-CRINI, La Revue Électronique Du Centre de Recherche Sur Les Identités*, 4, 1-20.
- Arnold, G. F. (1957). Stress in English words. *Lingua*, 6, 221-441.
- Arvaniti, A. (2020). The Phonetics of Prosody. In A. Arvaniti, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.411>
- Asher, J. J. (1969). The Total Physical Response Approach to Second Language Learning. *The Modern Language Journal*, 53(1), 3–17. <https://doi.org/10.2307/322091>
- Aslam, M., & Kak, A. A. (2007). *Introduction to English Phonetics and Phonology* (1st ed.). Foundation Books. <https://doi.org/10.1017/UPO9788175968653>
- Astésano, C. (2001). *Rythme et accentuation en français. Invariance et variabilité stylistique*. L'Harmattan.
- Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, 2, 89–195. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Attaheri, A., Choidealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., et al. (2022). Delta- and theta-band cortical tracking and phase amplitude coupling to sung speech by infants. *Neuroimage* 247, 118698. <https://doi.org/10.1016/j.neuroimage.2021.118698>
- Attardo, S., Wagner, M. M., & Urios-Aparisi, E. (Eds.). (2013). *Prosody and humor*. Benjamins.
- Atwood, M. (1996). *The Handmaid's Tale*. Vintage.
- Aubin, S. (2022). Didactique et méthodologie du rythme du français pour apprenants adultes : Un état des lieux et des perspectives. *Journal for Foreign Languages*, 14(1), 177–195. <https://doi.org/10.4312/vestnik.14.177-195>
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189–208. [https://doi.org/10.1016/S0021-9924\(03\)00019-4](https://doi.org/10.1016/S0021-9924(03)00019-4)
- Baddeley, A. (2021). Developing the Concept of Working Memory: The Role of Neuropsychology. *Archives of Clinical Neuropsychology*, 36(6), 861–873. <https://doi.org/10.1093/arclin/acab060>
- Baddeley, A., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). Academic Press.

- Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child-Adult Differences in Second-Language Phonological Learning: The Role of Cross-Language Similarity. *Language and Speech*, 51(4), 317–342. <https://doi.org/10.1177/0023830908099068>
- Baills, F., Alazard-Guiu, C., & Prieto, P. (2022). Embodied Prosodic Training Helps Improve Accentedness and Suprasegmental Accuracy. *Applied Linguistics*, 43(4), 776–804. <https://doi.org/10.1093/applin/amac010>
- Bandura, A. (1977a). *Social Learning Theory*. General Learning Press.
- Bandura, A. (1977b). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59(1), 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Baratta, A. (2016). Keeping it real or selling out: The effects of accent modification on personal identity. *Pragmatics and Society*, 7(2), 291–319. <https://doi.org/10.1075/ps.7.2.06bar>
- Baratta, A. (2018). “I Speak How I Speak:” A Discussion of Accent and Identity Within Teachers of ELT. In B. Yazan & N. Rudolph (Eds) *Criticality, Teacher Identity, and (In)equity in English Language Teaching. Educational Linguistics*, vol 35. Springer. <https://doi.org/10.1007/978-3-319-72920-6>
- Baratta, A. (2019). *World Englishes in English Language Teaching*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-13286-6>
- Barrett, D., & Heale, R. (2020). What are Delphi studies? *Evidence Based Nursing*, 23(3), 68–69. <https://doi.org/10.1136/ebnurs-2020-103303>
- Barbosa Pereira, C., u, X., Czaplik, M., Blazek, V., Venema, B., & Leonhardt, S. (2017). Estimation of breathing rate in thermal imaging videos: A pilot study on healthy human subjects. *Journal of Clinical Monitoring and Computing*, 31, 1241–1254. <https://doi.org/10.1007/s10877-016-9949-y>
- Baten, L., Beavan, A., Osborne, J., & Van Maele, J. (2013). WebCEF: An online collaboration tool for assessing foreign language proficiency. In Paolo M. Pumilia-Gnarini, E. Favaron, E. Pacetti, J. Bishop, & L. Guerra (Eds.), *Handbook of Research on Didactic Strategies and Technologies for Education: Incorporating Advancements*. IGI Global. <https://doi.org/10.4018/978-1-4666-2122-0>
- Bates, E; MacWhinney, B. (1982). Functionalist approaches to grammar. In Wanner, E. & Gleitman, L.R. (Eds.) *Language acquisition: the state of the art* (pp. 173-218). Cambridge University Press.

Bauvois, G. (2019). "France has never been and never will be a multicultural country" (Super)-diversity in Macron's France. *Siirtolaisuus - Migration*, 45(2), 8-10.

Beaumont J.G. (2008). *Introduction to Neuropsychology*. The Guilford Press.

Behra, S. (2019). Faut-il parler l'anglais depuis toujours pour bien l'enseigner ? *The Conversation*. <https://theconversation.com/faut-il-parler-langlais-depuis-toujours-pour-bien-lenseigner-115511>

Ben-David, B. M., Gal-Rosenblum, S., van Lieshout, P. H. H. M., & Shakuf, V. (2019). Age-Related Differences in the Perception of Emotion in Spoken Language: The Relative Roles of Prosody and Semantics. *Journal of Speech, Language, and Hearing Research*, 62(4S), 1188–1202. [https://doi.org/10.1044/2018\\_JSLHR-H-ASCC7-18-0166](https://doi.org/10.1044/2018_JSLHR-H-ASCC7-18-0166)

Benguerel, A-P. (1973). Correlélats physiologiques de l'accent en français. *Phonetica* 27, 21–35.

Benson, P. (2012). Learner-centered teaching. In A. Burns, & J. C. Richards (Eds.), *The Cambridge guide to pedagogy and practice in second language teaching* (pp. 30-37). Cambridge University Press.

Berry, C. (1991). *Voice and the Actor*. Wiley.

Bertrán, A.P. (1999). Prosodic Typology: On the Dichotomy between Stress-Timed and Syllable-Timed Languages. *Language Design Journal of Theoretical and Experimental Linguistics*, 103–131.

Bertin, J.C. (2000) Le système Learning Labs : une tentative pour allier recherche et développement. *Les Langues Modernes*, 3. 46-53.

Bertinetto, P.M., 1989. Reflections on the dichotomy "stress" vs. "syllable-timing". *Revue de Phonétique Appliquée*, 91-93, 99-130.

Bertolero, M. A., & Bassett, D. S. (2020). On the nature of explanations offered by network science: a perspective from and for practicing neuroscientists. *Topics in Cognitive Science*, 12(4), 1272–1293. <https://doi.org/10.1111/tops.12504>

Best, C. T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). MIT Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Learning & Language Teaching* (Vol. 17, pp. 13–34). John Benjamins. <https://doi.org/10.1075/llt.17.07bes>

- Billières, M., Alazard-Guiu, C., Berdoulat, H., Billerey, B., Boureux, M., Briet, G., Bussutil, C., Collige Neuenschwander, V., Marijanovic, V., Palusci, S., & Rassart, E. (2018). Mooc « Pratiques de l'enseignement de la prononciation en FLE ». [www.fun-mooc.fr/courses/course-v1:univ-toulouse+101016+session01/about](http://www.fun-mooc.fr/courses/course-v1:univ-toulouse+101016+session01/about)
- Birdsong, D. (2018). Plasticity, Variability and Age in Second Language Acquisition and Bilingualism. *Frontiers in Psychology, 9*, 81. <https://doi.org/10.3389/fpsyg.2018.00081>
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company.
- Bloomfield, L. (1933). *Language*. Holt, Rinehart & Winston.
- Bolinger, D. L. (1958). A Theory of Pitch Accent in English. *Word, 14*(2–3), 109–149. <https://doi.org/10.1080/00437956.1958.11659660>
- Bongaerts, T., Mennen, S., & Slik, F. van der. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica, 54*(2), 298–308. <https://doi.org/10.1111/1467-9582.00069>
- Bordal, G. (2012). A phonological study of French spoken by multilingual speakers from Bangui, the capital of the Central African Republic. In R. Gess, C. Lyche, & T. Meisenburg (Eds.), *Phonological variation in French: Illustrations from three continents* (pp. 23–43). Benjamins. <https://doi.org/10.1075/silv.11.03bor>
- Borrell, A., Salsignac J., Importance de la prosodie en didactique des langues (application au FLE). In R. Renard (Ed.), *Apprentissage d'une langue étrangère/seconde. Vol. 2* (pp. 163-182). De Boeck Supérieur <https://doi.org/10.3917/dbu.renar.2002.01.0163>
- Boula de Mareüil, P. & Vieru-Dimulescu, B. (2006). The Contribution of Prosody to the Perception of Foreign Accent. *Phonetica, 63*(4), 247–267. <https://doi.org/10.1159/000097308>
- Bourrain, S., Dupayage, V., Walgenwitz, G., & Corvaisier, M. (2021). *Le corps au coeur des apprentissages : Faire cours, un peu, autrement 97 activités et 8 éclairages théoriques*. Presses Universitaires de Grenoble.
- Boureux, M., Alazard Guiu, C., & Billières, M. (2021). Former à l'enseignement de la prononciation du FLE par la méthode verbo-tonale. Du présentiel au distanciel. In J. Sauvage (Ed.), *Didactique de la phonétique du français : Et maintenant ? (Le Langage et l'Homme. Revue de didactique du français)* (pp. 137–156), L'Harmattan.

- Bower, K. (2019). Explaining motivation in language learning: A framework for evaluation and research. *The Language Learning Journal*, 47(5), 558–574. <https://doi.org/10.1080/09571736.2017.1321035>
- Braud, V., Millot, P., Sarré, C., & Wozniak, S. (2015). « You say you want a revolution... » Contribution à la réflexion pour une politique des langues adaptée au secteur LANSAD. *Recherche et pratiques pédagogiques en langues de spécialité - Cahiers de l'APLIUT*, 34(1), 46-66. <https://doi.org/10.4000/apliut.5020>
- Braud, V., Millot, P., Sarré, C., & Wozniak, S. (2016). *Quelles conceptions de la maîtrise de l'anglais en contexte professionnel ? Vers une définition de la « compétence en anglais de spécialité »*. *Mélanges CRAPEL, Varia*, 37, 13-44.
- Braun, B., Lemhöfer, K., & Mani, N. (2011). Perceiving unstressed vowels in foreign-accented English. *The Journal of the Acoustical Society of America*, 129(1), 376–387. <https://doi.org/10.1121/1.3500688>
- Brazil, D. (1975). *Discourse intonation I. Discourse Analysis Monograph 1*. University of Birmingham English Language Research.
- Brazil, D. (1978). *Discourse intonation II. Discourse Analysis Monograph 2*. University of Birmingham English Language Research.
- Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge University Press.
- Briere, E. J. (1966). An Investigation of Phonological Interference. *Language*, 42(4), 768. <https://doi.org/10.2307/411832>
- Bronfenbrenner, U. (1974). Developmental research, public policy, and the ecology of childhood. *Child development*, 45(1), 1-5. <https://doi.org/10.2307/1127743>
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32(7), 513–531. <https://doi.org/10.1037/0003-066X.32.7.513>
- Bronfenbrenner, U. (1994). Ecological models of human development. In M. Gauvain, & M. Cole (Eds.), *Readings on the development of children* (2nd ed., pp. 37–43). Freeman.
- Brown, A. (1988). Functional Load and the Teaching of Pronunciation. *TESOL Quarterly*, 22(4), 593. <https://doi.org/10.2307/3587258>
- Brumfit, C.J. (2001). *Individual Freedom in Language Teaching: Helping Learners to Develop a Dialect of their Own*. Oxford University Press.
- Bruner, J. (1960). *The process of education*. Harvard University Press.

- Buisson-Fenet, H. (2014). Enseigner les langues en Europe : Expertise communautaire ou expertise d'État ? : *Carrefours de l'éducation*, 37(1), 79–94. <https://doi.org/10.3917/cdle.037.0079>
- Burke, D. M., & Shafto, M. A. (2004). Aging and Language Production. *Current Directions in Psychological Science*, 13(1), 21–24. <https://doi.org/10.1111/j.0963-7214.2004.01301006.x>
- Burri, M., & Baker, A. (2021). 'I Feel ... Slightly out of Touch': A Longitudinal Study of Teachers Learning to Teach English Pronunciation over a Six-Year Period. *Applied Linguistics*, 1–19. <https://doi.org/10.1093/applin/amab009>
- Campfield, D. E., & Murphy, V. A. (2013). The influence of prosodic input in the second language classroom: Does it stimulate child acquisition of word order and function words? *The Language Learning Journal*, 45(1), 81–99. <https://doi.org/10.1080/09571736.2013.807864>
- Capliez, M. (2016). *Acquisition and learning of English phonology by French speakers: On the roles of segments and suprasegments* [PhD thesis]. Lille III University.
- Cardoso, W. (2007). The development of sC onset clusters in interlanguage: markedness vs. frequency effects. In R. Slabakova, J. Rothman, P. Kempchinsky, and E. Gavruseva (Eds.). *Proceedings of the 9th generative approaches to second language acquisition conference* (pp. 15–29). Cascadilla Press.
- Carpenter, A. C. (2015). Phonetic Training Significantly Mitigates the Stress 'Deafness' of French Speakers. *International Journal of Linguistics*, 7(3), 94–108. <https://doi.org/10.5296/ijl.v7i3.7661>
- Carrera-Sabaté, J. (2021) Body to Speech and Back: Considerations on Embodied Pronunciation. Proc. 3rd International Symposium on Applied Phonetics (ISAPh 2021), 1-8, <https://doi.org/10.21437/ISAPh.2021-1>
- Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-analytic Studies*. Cambridge University Press.
- Catford, J. (1950) Intelligibility. *English Language Teaching Journal*, 1(1), 7-15.
- Catford, J. C. (1987). Phonetics and the teaching of pronunciation: A systemic description of English phonology. In J. Morley Current perspectives on pronunciation: Practices (ed.), anchored in theory. Washington, DC: TESOL, 87–100.
- Cauldwell, R. (1996). Stress-timing: Observations, beliefs, and evidence. *Eger Journal of English Studies*, 1, 33–48.
- Cauldwell, R.T. (2013): Phonology for Listening: Teaching the stream of speech. Birmingham: Speech in Action, *System* 42, 296-297.

- Cave, C., Guaitella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). About the relationship between eyebrow movements and Fo variations. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, 4, 2175–2178. <https://doi.org/10.1109/ICSLP.1996.607235>
- Cazade, A. (1999). De l'usage des courbes sonores et autres supports graphiques pour aider l'apprenant en langues. *Alsic, Vol. 2, n° 2*, 3–32. <https://doi.org/10.4000/alsic.1623>
- Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge University Press.
- Chambers. (1993). Accent. In *The Chambers Dictionary*. Chambers, 8.
- Chambers. (1993). Intelligible. In *The Chambers Dictionary*. Chambers, 870.
- Chambers. (1993). Voice. In *The Chambers Dictionary*. Chambers, 1946.
- Chan, M. (2018). Embodied Pronunciation Learning: Research and Practice. *The CATESOL Journal*, 30(1), 47–68.
- Champagne-Muzar, C., Schneiderman, E. and J. S. Bourdages. 1993. Second Language Accent: The Role of the Pedagogical Environment. *IRAL: International Review of Applied Linguistics in Language Teaching* 31, 143.
- Chaplier, C. (2013). Des cours de sciences en anglais à l'EMILE : État des lieux, réflexion et recommandations : Cas de l'Université Paul Sabatier. *Recherche et pratiques pédagogiques en langues de spécialité - Cahiers de l'APLIUT*, 32(3), 57–79. <https://doi.org/10.4000/apliut.3867>
- Chen, H. C., & Tian, J. (2020). The Effects of Explicit Rule and Acoustic-perceptual Instructions on Chinese ESL Learners' Prosodic Acquisition of English Lexical Stress. *Speech Prosody 2020*, 833–837. <https://doi.org/10.21437/SpeechProsody.2020-170>
- Chini, D. (2001). Médiation(s) : Quelques remarques sur un terme pluriel. *ASP*, 31–33, 133–142. <https://doi.org/10.4000/asp.1901>
- Chomsky, N., Halle, M. (1968) *The Sound Pattern of English*. Harper and Row.
- Chomsky, N. (1969). *Aspects of the Theory of Syntax*. MIT Press.
- Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, 4, 1–11. <https://doi.org/10.3389/fpsyg.2013.00874>

- Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping Lexical and Syntactic Acquisition. *Language and Speech*, 51(1–2), 61–75. <https://doi.org/10.1177/00238309080510010501>
- Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Benjamins.
- Chung, D. (2018). Language Neuromechanics: The Human Biological-Language Evolution. *Journal of Behavioral and Brain Science*, 08(08), 447–472. <https://doi.org/10.4236/jbbs.2018.88028>
- Clark, A. (1999). Where brain, body, and world collide. *Cognitive Systems Research* 1 (1), 5-17. [https://doi.org/10.1016/S1389-0417\(99\)00002-9](https://doi.org/10.1016/S1389-0417(99)00002-9)
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647–3658. <https://doi.org/10.1121/1.1815131>
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210. <https://doi.org/10.1007/BF01320076>
- Cleland, J., Peppé, S.J.E. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, 38(4), 325-50. <https://doi.org/10.1080/1368282031000154204>
- Coates, J. (2015). *Women, men and language: a sociolinguistic account of gender differences in language*. Routledge.
- Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning: A systematic and critical review*. Learning and Skills Research Centre.
- Collen, I. (2020). *Language Trends 2020 Language teaching in primary and secondary schools in England*. The British Council.
- Cooke, R. (1993) Reducing Word Stress Errors: Time Restricted Help for ESP Students, *Asp*, 2, 164-175.
- Cooper, Christopher. (1688). *The English teacher, or, The discovery of the art of teaching and learning the English tongue fitted for the use of schools and necessary for all those that desire to read, write, or speak our tongue with ease and understanding*. Printed by John Richardson for George Coniers
- Cooper, W. E., & Eady, S. J. (1986). Metrical phonology in speech production. *Journal of Memory and Language*, 25(3), 369–384. [https://doi.org/10.1016/0749-596X\(86\)90007-0](https://doi.org/10.1016/0749-596X(86)90007-0)
- Coulange, S., & Rossato, S. (2020). Proximité rythmique entre apprenants et natifs du français Évaluation d’une métrique basée sur le CEFC. 6e conférence conjointe Journées d’Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des



Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 1. Journées d'Études sur la Parole, 2020, Nancy, France, 118-126.

Council of Europe. (2001). *A Common European Framework of Reference for learning, teaching and assessment*. Cambridge University Press.

Council of Europe. (2020). *A Common European Framework of Reference for learning, teaching and assessment. Companion Volume*. Council of Europe Publishing.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect: An Australian Journal of TESOL*, 21(1), 46–66.

Couper, G. F. (2009). *Teaching and Learning L2 Pronunciation*: [PhD thesis]. University of New England.

Couper, G. (2019). Teachers' cognitions of corrective feedback on pronunciation: Their beliefs, perceptions and practices. *System*, 84, 41–52. <https://doi.org/10.1016/j.system.2019.04.003>

Couper, G. (2022a). A conceptual approach to teaching L2 pronunciation: Perception of word stress. *TESL-EJ* 26(1), 1-20. <https://doi.org/10.55593/ej.26101a6>

Couper, G. (2022b). Teaching and testing perception of word stress: many shades of perception. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharine's, ON. <https://doi.org/10.31274/psllt.13266>

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Edward Arnold.

Crawford Camiciottoli, B., & Campoy-Cubillo, M. C. (2018). Introduction: The nexus of multimodality, multimodal literacy, and English language teaching in research and practice in higher education settings. *System*, 77, 1–9. <https://doi.org/10.1016/j.system.2018.03.005>

Cruttenden, A. (1986). *Intonation* (First edition). Cambridge University Press.

Cruttenden, A. (1997). *Intonation* (Second edition). Cambridge University Press.

Cruttenden, A., & Gimson, A. C. (2014). *Gimson's pronunciation of English* (Eighth Edition). Routledge.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). *Prosody in the Comprehension of Spoken Language: A Literature Review*. *Language and Speech*, 40 (2), 62, 141-201. <https://doi.org/10.1177/002383099704000203>

Cutler, C. (2014). Accentedness, "passing" and crossing. In J. M. Levis & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 145–167). DeGruyter. <https://doi.org/10.1515/9781614511762.145>

- Crystal, David (1969). *Prosodic Systems and Intonation in English*. Cambridge University Press.
- Crystal, D. (1997). *English as a global language*, (Second edition). Cambridge University Press.
- Crystal, D. (2008). *A dictionary of linguistics and phonetics*. (Sixth edition). Blackwell Publishing.
- Cyrulnik, B. (2022). Language, the Brain, and Relating. In S. Greaves & M. De Mattia-Viviès (Eds.), *Language Learning and the Mother Tongue: Multidisciplinary Perspectives* (pp. 171-186). Cambridge University Press. <http://doi.org/10.1017/9781009029124.013>
- Cuq, J.-P. (Ed.). (2003). *Dictionnaire de didactique du français langue étrangère et seconde*. CLE International.
- Cuq, J.-P. 2005. « Trente ans d'évolution de la didactique des langues étrangères et secondes en France vues à travers deux dictionnaires *La Revue de l'AQEFLS*, 25(2), pp. 45-61.
- Ćwiek, A., & Fuchs, S. (2019). Iconic Prosody is Rooted in Sensori-Motor Properties: Fundamental Frequency and the Vertical Space. *CogSci 2019: 41st Annual Meeting of the Cognitive Science Society*, 1–8.
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. John Murray.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51–62. [https://doi.org/10.1016/S0095-4470\(19\)30776-4](https://doi.org/10.1016/S0095-4470(19)30776-4)
- Davis, B.L., MacNeilage, P.F. (1995). The articulatory basis of babbling. *Journal of speech and hearing research*, 38 6, 1199-211. <https://doi.org/10.1044/jshr.3806.1199>
- de Bot, K., & Mailfert, K. (1982). The Teaching of Intonation: Fundamental Research and Classroom Applications. *TESOL Quarterly*, 16(1), 71. <https://doi.org/10.2307/3586564>
- de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced Language and Speech practice behavior. *Language and Speech*, 26(4), 331-350. <https://doi.org/10.1177/00238309830260040>
- Deci, E.L. & Ryan, R.M. (1980). Self-determination theory: When mind mediates behavior. *The Journal of Mind and Behavior*, 1(1), 33–43.
- Delattre, P. (1938) L'accent final en français: accent d'intensité, accent de hauteur, accent de durée, *The French Review*, 12(2), 141-145. <https://doi.org/10.1515/iral.1963.1.1.193>

- Delattre, P. (1963). Comparing the Phonetic Features of English, French, German and Spanish. *IRAL - International Review of Applied Linguistics in Language Teaching*, 1(1). <https://doi.org/10.1515/iral.1963.1.1.193>
- Delattre, P. (1966a). Les dix intonations du français. *The French Review*, 40, 1-14
- Delattre P. (1966b). *Studies in French and Comparative Phonetics*. Mouton.
- Derwing, T. M. (1990). Speech Rate Is No Simple matter: Rate Adjustment and NS–NNS Communicative Success. *Studies in Second Language Acquisition*, 12(3), 303–313. <https://doi.org/10.1017/S0272263100009189>
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility and comprehensibility: Evidence from FourL1s. *Studies in Second Language Acquisition*, 19(1), 1–16. <https://doi.org/10.1017/S0272263197001010>
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1997). Pronunciation instruction for ‘fossilized’ learners. Can it help? *Applied Language Learning*, 8(2), 217–235.
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in Favor of a Broad Framework for Pronunciation Instruction. *Language Learning*, 48(3), 393–410. <https://doi.org/10.1111/0023-8333.00047>
- Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29(2), 217–230. <https://doi.org/10.1006/jpho.2001.0138>
- Detey, S., & Racine, I. (2012). Les apprenants de français face aux normes de prononciation : Quelle(s) entrée(s) pour quelle(s) sortie(s) ? *Revue française de linguistique appliquée*, Vol. XVII(1), 81–96. <https://doi.org/10.3917/rfla.171.0081>
- Dewaele, J.-M. (2017). Why the Dichotomy ‘L1 Versus LX User’ is Better than ‘Native Versus Non-native Speaker’. *Applied Linguistics*, 236–240. <https://doi.org/10.1093/applin/amw055>
- Díaz-Posada, L.-E., Varela-Londoño, S., & Rodríguez-Burgos, L.-P. (2017). Multiple Intelligences and Curriculum Implementation: Progress, Trends and Opportunities. *Revista de Psicodidáctica*, 22(1), 69–83.
- Di Cristo, Albert. (1998). Intonation in French. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 88–103). Cambridge University Press.
- Di Cristo, A. (2004). La prosodie au carrefour de la phonétique, de la phonologie et de l’articulation formes-fonctions. *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, 23, 67–211.

Didelot, M., Racine, I., Zay, F., & Prikhodkine, A. (2019). Enseignement et évaluation de la prononciation aujourd'hui : L'intelligibilité comme enjeu. *Recherches en didactique des langues et des cultures*, 16(1). <https://doi.org/10.4000/rdlc.4333>

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation☆. *Journal of Memory and Language*, 63(3), 274–294. <https://doi.org/10.1016/j.jml.2010.06.003>

Dohen, M., & Løevenbruck, H. (2009). Interaction of Audition and Vision for the Perception of Prosodic Contrastive Focus. *Language and Speech*, 52(2–3), 177–206. <https://doi.org/10.1177/0023830909103166>

Dörnyei, Z. (2007). *Research methods in Applied Linguistics. Quantitative, qualitative and mixed methodologies*. Oxford University Press.

Di Meo, A. (2012). Interlanguage prosody. In A. De Meo & M. Pettorino (Eds.), *Prosodic and Rhythmic Aspects of L2 Acquisition: The Case of Italian*. Cambridge: Cambridge Scholars Publishing.

Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M., & Hon, H.-W. (2019). Unified Language Model Pre-training for Natural Language Understanding and Generation. *NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 13063–13075.

Dragojevic, M., & Giles, H. (2016). I Don't Like You Because You're Hard to Understand: The Role of Processing Fluency in the Language Attitudes Process: Processing Fluency and Language Attitudes. *Human Communication Research*, 42(3), 396–420. <https://doi.org/10.1111/hcre.12079>

Dressen-Hammouda, D., & Wigham, C. R. (2022). Evaluating multimodal literacy: Academic and professional interactions around student-produced instructional video tutorials. *System*, 105, 102727. <https://doi.org/10.1016/j.system.2022.102727>

Ducancel, G. (1992). Enseignement des langues vivantes et enseignement du Français à l'école: Des interactions et des recherches à construire. *Repères*, 6(1), 3–11. <https://doi.org/10.3406/reper.1992.2061>

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A Destressing "Deafness" in French? *Journal of Memory and Language*, 36(3), 406–421. <https://doi.org/10.1006/jmla.1996.2500>

Dupoux, E., & Peperkamp, S. (2001). A robust method to study stress "deafness". *Journal of the Acoustic Society of America*, 110(3), 1606–1618. <https://doi.org/10.1121/1.1380437>

- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, 106(2), 682–706. <https://doi.org/10.1016/j.cognition.2007.04.001>
- Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27(2), 315–330. <https://doi.org/10.1111/j.1467-1770.1977.tb00124.x>
- Eckman, F. R. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition*, 26(04). <https://doi.org/10.1017/S027226310404001X>
- Eckman, F. R. (2008). 4. Typological markedness and second language phonology. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Studies in Bilingualism* (Vol. 36, pp. 95–115). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.36.06eck>
- Edensor, K. (2008). French comprehension of English regional accents. *TIPA. Travaux Interdisciplinaires Sur La Parole et Le Langage*, 27, 51–62. <https://doi.org/10.4000/tipa.281>
- Edensor, K. (Forthcoming). The Impact of Day-to- Day Use of Multimedia by L2 Learners of English on the Comprehension of Regional Varieties.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.
- Engelkamp, J., & Zimmer, H. D. (1989). Memory for action events: A new field of research. *Psychological Research*, 51(4), 153–157. <https://doi.org/10.1007/BF00309142>
- Engelkamp, J. (1998). *Memory for actions*. Psychology Press (Taylor & Francis).
- Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review*, 100(3), 363–406. <https://doi.org/10.1037/0033-295X.87.3.215>
- Ericsson, K. A., & Harwell, K. W. (2019). Deliberate Practice and Proposed Limits on the Effects of Practice on the Acquisition of Expert Performance: Why the Original Definition Matters and Recommendations for Future Research. *Frontiers in Psychology*, 10, 2396. <https://doi.org/10.3389/fpsyg.2019.02396>
- Eschenauer, S. (2021). Oser la performance théâtrale plurilingue à l'école pour une éducation au développement durable. *Questions vives recherches en éducation*, N° 35. <https://doi.org/10.4000/questionsvives.5641>

Eschenauer, S., Tellier, M., & Zappa, A. (2022). Encorporer les langues vivantes : Reconnaître la place du corps pour enseigner et pour apprendre. *TIPA. Travaux interdisciplinaires sur la parole et le langage*, 38. <https://doi.org/10.4000/tipa.4790>

European Commission. (2012a). *Europeans and their languages* (Conducted by TNS Opinion & Social at the request of Directorate-General Education and Culture, Directorate-General for Translation and Directorate-General for Interpretation. Survey co-ordinated by the European Commission, Directorate-General for Communication (DG COMM “Research and Speechwriting” Unit)). European Union Publications Office.

European Commission. Directorate General for Education and Culture. (2012b). *First European survey on language competences: final report*. European Union Publications Office.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, 73(6), 2608–2611. <https://doi.org/10.1152/jn.1995.73.6.2608>

Fant, G. (1960) *Acoustic theory of speech production*. Mouton.

Farrús, M., & Codina-Filbà, J. (2020). *Combining Prosodic, Voice Quality and Lexical Features to Automatically Detect Alzheimer’s Disease*. *arXiv [Preprint]*. Retrieved 10 May, 2023 from <https://arxiv.org/abs/2011.09272>

Faure, G. (1948). *Manuel Pratique d’Anglais Parlé*. Librairie Hachette.

Faure G. (1971a). Fundamental tendencies of French phoneticism and their pedagogic implications in teaching of the prosodic structures of French to foreign students. *Etudes de Linguistique Appliquée*, 3, 1-14.

Faure, G. (1971b). La description phonologique des systèmes prosodiques. *STUF - Language Typology and Universals*, 24(1–6), 347–359. <https://doi.org/10.1524/stuf.1971.24.16.347>

Fayer, J. M., & Krasinski, E. (1987). Native and Nonnative Judgments of Intelligibility and Irritation. *Language Learning*, 37(3), 313–326. <https://doi.org/10.1111/j.1467-1770.1987.tb00573.x>

Fagyal, Z. (2010). *Accents de banlieue : aspects prosodiques du français populaire en contact avec les langues de l’immigration*. L’Harmattan.

Faure, G., Albert Di Cristo, A. (1973). *Phonétique générale et phonétique descriptive du français*. University of Michigan Press.

- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, 37(1), 1–9. <https://doi.org/10.3758/MC.37.1.1>
- Fenclová, M. (2014). Langue seconde, langue étrangère et aspects cognitifs : *Éla. Études de linguistique appliquée*, n° 174(2), 147–155. <https://doi.org/10.3917/ela.174.0147>
- Fernald, A., & Mazzie, C. (1991). Prosody and Focus in Speech to Infants and Adults. *Developmental Psychology*, 27(2), 209–221. <https://doi.org/10.1037/0012-1649.27.2.209>
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501. <https://doi.org/10.1017/S0305000900010679>
- Ferreira, A. M. A., Barbosa, L. M. de A., & dos Reis, M. da G. M. (2010). Dictionnaire de didactique des langues de R. Galisson et D. Coste, et Dictionnaire de didactique du français langue étrangère et seconde de J.P. Cuq : Quelles organisations ? *Synergies Brésil*, 8, 49–56.
- Fisher, J., & Kayes, G. (2016). This is a voice: 99 exercises to train, project and harness the power of your voice. Wellcome Collection.
- Firth, S. (1992). Pronunciation syllabus design: a question of focus. In P. Avery & S. Ehrlich (Eds.), *Teaching American-English pronunciation* (pp. 173-183). Oxford University Press.
- Fiske, J. (2023). The Genesis of Language. *The North American Review*, 109(225), 305–367.
- Fitch, W. T. (2011). The biology and evolution of rhythm: Unravelling a paradox. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 73–95). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199553426.003.0009>
- Fitch, W. T. (2013). Musical Protolanguage: Darwin's Theory of Language Evolution Revisited. In J. J. Bolhuis & M. Everaert (Eds.), *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*. MIT Press.
- Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, 44(3), 329–344. <https://doi.org/10.1037/h0057198>
- Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, 76(3), 692–707. <https://doi.org/10.1121/1.391256>
- Flege, J. E. (1995). Theory, Findings, and Problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, J. E. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*, 21(5), 919–920. <https://doi.org/10.1017/S136672891800010X>

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model. In R. Wayland (Ed.), *Second Language Speech Learning* (pp. 3–83). Cambridge University Press. <https://doi.org/10.1017/9781108886901.002>

Flemming, E. (2009). The Phonetics of Schwa Vowels. In D. Minkova (Ed.), *Phonological Weakness in English* (pp. 78–95). Palgrave Macmillan UK. [https://doi.org/10.1007/978-0-230-29686-2\\_5](https://doi.org/10.1007/978-0-230-29686-2_5)

Foote, J. A., & Thomson, R. I. (2021). Speech language pathologists' beliefs and knowledge-base for providing pronunciation instruction: A critical survey. *Journal of Second Language Pronunciation*, 7(2), 240–264. <https://doi.org/10.1075/jslp.20031.foo>

Fónagy I. (1979), L'accent français : accent probabilitaire, in I. Fónagy & P. Léon (Eds), *L'Accent en français contemporain (Studia Phonetica 15)* 123-233. Didier.

Fonseca-Mora, M. C., & Machancoses, F. H. (2016). 16 Music and Language Learning: Emotions and Engaging Memory Pathways. In P. D. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive Psychology in SLA* (pp. 359–373). Multilingual Matters. <https://doi.org/10.21832/9781783095360-017>

Foucart, A., & Brouwer, S. (2021). Is There a Foreign Accent Effect on Moral Judgment? *Brain Sciences*, 11(12), 1631. <https://doi.org/10.3390/brainsci11121631>

Fougeron, C., & Smith, C. (1993). French. *Journal of the International Phonetic Association*, 23(2), 73-  
<https://doi.org/10.1017/S0025100300004874>

Fouz-González J (2015) Trends and directions in computer-assisted pronunciation training. In: Mompean JA, Fouz-González J (eds) *Investigating English Pronunciation: Trends and Directions*. Palgrave Macmillan, 314–342. [https://doi.org/10.1057/9781137509437\\_14](https://doi.org/10.1057/9781137509437_14)

Fouz-González, J. (2017). Pronunciation instruction through Twitter: The case of commonly mispronounced words. *Computer Assisted Language Learning*, 30(7), 631–663. <https://doi.org/10.1080/09588221.2017.1340309>

Fowler, C. (2016). Speech perception as a perceptuo-motor skill. In G. Hickok & S. Small (Eds.) *Neurobiology of language* (pp. 175-184) Elsevier.

Fraser, H. (2006). Helping teachers help students with pronunciation: A cognitive approach. *Prospect: An Australian Journal of TESOL*, 21(1), 80–96.



- Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 349–366. <https://doi.org/10.1037/0096-1523.28.2.349>
- Fries, C. C. (1945). *Teaching and Learning English as a Foreign Language*. Ann Arbor: University of Michigan Press.
- Frost, D. (2002). Seeing is believing : L’oscillogramme, l’auto-apprentissage et l’accentuation de mots. *Les Cahiers de l’APLIUT*, 21(3), 21–31.
- Frost, D. (2003). La phonétique pour les vaches espagnoles (fiche pédagogique pour présenter et enseigner les sons de l’anglais). *Les Cahiers de l’APLIUT*, 22(3), 39-52.
- Frost, D. (2008). *The Stress Site*. L’accentuation et la compréhension de l’anglais oral : le distanciel peut-il remplacer le présentiel ? *Asp*, 53-54, 111-127. <https://doi.org/10.4000/asp.382>
- Frost, D. (2011). Stress cues in English and French: A perceptual study. *Journal of the International Phonetic Association*, 41(01), 67–84. <https://doi.org/10.1017/S0025100310000253>
- Frost, D. (2021). Prosodie, intelligibilité et compréhensibilité : L’évaluation de la prononciation lors d’un stage court. *Les Langues Modernes*, 3(2020), 76–90.
- Frost, D., & Guy, R. (2016). L’innovation est le ton qui fait la chanson dans le secteur LANSAD : musique et prosodie dans le projet Innovalangues. *Recherche et Pratiques Pédagogiques En Langues de Spécialité – Cahiers de l’APLIUT*, 35(1), 1–21. <https://doi.org/10.4000/apliut.5526>
- Frost, D., & Henderson, A. (2013). Les résultats du sondage EPTiES (English Pronunciation Teaching in Europe Survey) : L’enseignement de la prononciation dans plusieurs pays européens vu par les enseignants. *Recherche et Pratiques Pédagogiques En Langues de Spécialité – Cahiers de l’APLIUT*, 32(1), 92–113. <https://doi.org/10.4000/apliut.3586>
- Frost, D., Henderson, A., Haidar, L. A., & Jordan, W. (2021). Rationale and design of a study of foreign accented academic English. In A. Kirkova-Naskova, A. Henderson, & J. Fouz-González (Eds.), *English Pronunciation Instruction. Research-based insights* (Vol. 18, pp. 197–222). John Benjamins.
- Frost, D., & O’Donnell, J. (2013). Combatting the “can’t do mentality”: Expert, peer & self-assessment in a French university context (short paper). *Proceedings of the 2nd International Conference “Language Testing in Europe: Time for a New Framework?”*, 104–109.

- Frost, D., & O'Donnell, J. (2015). Success: B2 or not B2, that is the question (the ELLO project—Etude Longitudinale sur la Langue Orale). *Recherche et Pratiques Pédagogiques En Langues de Spécialité – Cahiers de l'APLIUT*, 34(2), pagination en cours. <https://doi.org/10.4000/apliut.5195>
- Frost, D., & O'Donnell, J. (2018). Evaluating the essentials: The place of prosody in oral production. In J. Volín (Ed.), *The Pronunciation of English by Speakers of Other Languages* (pp. 228–259). Cambridge Scholars Publishing.
- Frost, D., & Picavet, F. (2014). Putting prosody first – some practical solutions to a perennial problem: The Innovalangues project. *Research in Language*, 12(3), 1–11.
- Fry, D.B. (1947). The frequency of occurrence of speech sounds in Southern English. *Archives néerlandaises. de Phonétique. expérimentale*, 20, 103-6.
- Fry, D.B. (1955) Duration and Intensity as Physical Cues of Linguistic Stress, *Journal of the Acoustic Society of America*, 27(4), 765-768.
- Fry, D.B. (1958) Experiments in the Perception of Stress. *Language and Speech*, 1(2), 126-152.
- Fry, D.B. (1965) The Dependence of Stress Judgements on Vowel Formant Structure. In E. Zwirner et W. Bethge (Eds.) *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 306-311). Karger.
- Fujii, S., & Wan, C. Y. (2014). The Role of Rhythm in Speech and Language Rehabilitation: The SEP Hypothesis. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00777>
- Gabriel, C., & Kireva, E. (2014). Prosodic transfer in learner and contact varieties: Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition*, 36(2), 257–281. <https://doi.org/10.1017/S0272263113000740>
- Gabriel, C., Feldhausen, I., Pešková, A., Colantoni, L., Lee, S., Arana, V., & Labastía, L. (2010). Argentinian Spanish intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of intonation of the Spanish language* (pp. 285–317). Lincom. <https://doi.org/10.1515/opli-2017-0023>
- Galisson, R., Coste, D. (Eds.). (1976). *Dictionnaire de didactique des langues*. Hachette.
- Gaonac'h, D. (1987). *Théories d'apprentissage et acquisition d'une langue étrangère*. Hatier.
- Gaonac'h, D (Ed.) (1990). *Acquisition et utilisation d'une langue étrangère : l'approche cognitive*. Hachette.

- Gardner, H. (1983). *Frames of mind. The theory of multiple intelligences*. Basic Books.
- Gardner, H. (1995). Reflections on Multiple Intelligences: Myths and Messages. *Phi Delta Kappan*, 77(3), 200-203, 206-209.
- Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, 34(1), 65–87. <https://doi.org/10.1111/j.1467-1770.1984.tb00996.x>
- Gast, V. (2023). The Temporal Alignment of Speech-Accompanying Eyebrow Movement and Voice Pitch: A Study Based on Late Night Show Interviews. *Behavioral Sciences*, 13(1), 52. <https://doi.org/10.3390/bs13010052>
- Gathercole, S. (2008). Working Memory. In J. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 33–51). Elsevier.
- Gattegno, C. (1963). *Teaching Foreign Languages in Schools: The Silent Way*. Educational Explorers Limited.
- Genesee, F. (2015). Myths about early childhood bilingualism. *Canadian Psychology / Psychologie Canadienne*, 56(1), 6–15. <https://doi.org/10.1037/a0038599>
- Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, 4(1), 1490–1496. <https://doi.org/10.1038/ncomms2430age>
- Gervain, J., Christophe, A., Mazuka, R. (2021). Prosodic Bootstrapping, in C. Gussenhoven, and A. Chen (Eds), *The Oxford Handbook of Language Prosody*, Oxford University Press (pp. 563–573) <https://doi.org/10.1093/oxfordhb/9780198832232.013.36>
- Gibbon, D. (2018). The Future of Prosody: It's about Time. *Speech Prosody 2018*, 1–9. <https://doi.org/10.21437/SpeechProsody.2018-1>
- Gilbert, J. (1980). Prosodic development: Some pilot studies. *Los Angeles Second Language Acquisition Research Forum (Second, Los Angeles, California, October 6-8, 1978)*, 110–117.
- Gilbert, J. (2010). Pronunciation as orphan: what can be done? *SpeakOut: The newsletter of the IATEFL pronunciation special interest group*, 43, 3-7.
- Gilbert, J. (2014). Myth 4: intonation is hard to teach. In L. Grant (Ed.), *Pronunciation Myths: Applying Second Language Research to Classroom Teaching* (pp. 107–136). University of Michigan Press.
- Gillesen, C. (2021). *Rapport du jury, Agrégation externe, Langues vivantes étrangères, option anglais*. Ministère de l'éducation nationale, de la jeunesse et des sports. Retrieved 10 May, 2023 from

[https://media.devenirensignant.gouv.fr/file/agregation\\_externer/56/0/rj-2021-agregation-externe-lve-anglais\\_1419560.pdf](https://media.devenirensignant.gouv.fr/file/agregation_externer/56/0/rj-2021-agregation-externe-lve-anglais_1419560.pdf)

Gilquin, G., Bestgen, Y., & Granger, S. (2022). Assessing EFL Speech: A Teacher-Focused Perspective. *Journal of Second Language Teaching and Research*, 9.

Ginésy, M. (1989). *Exercices de phonétique*. Nathan.

Ginésy, M. (1995). *Mémento de phonétique anglaise*. Nathan.

Ginet, A (Ed.) (1997). *Du laboratoire de langues à la salle de cours multi-médias : de la recherche à la mise en pratique*. Nathan.

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3-48). Cambridge, England: Cambridge University Press.

Gluszek, A., Newheiser, A.-K., & Dovidio, J. F. (2011). Social Psychological Orientations and Accent Strength. *Journal of Language and Social Psychology*, 30(1), 28–45. <https://doi.org/10.1177/0261927X10387100>

Gogus, A. (2012). Bloom's Taxonomy of Learning Objectives. In N. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 469–473). Springer. [https://doi.org/10.1007/978-1-4419-1428-6\\_141](https://doi.org/10.1007/978-1-4419-1428-6_141)

Goswami, U. (2019). Speech rhythm and language acquisition: An amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*, 1453(1), 67–78. <https://doi.org/10.1111/nyas.14137>

Goswami, U. (2022). Language acquisition and speech rhythm patterns: An auditory neuroscience perspective. *Royal Society Open Science*, 9(7), 211855. <https://doi.org/10.1098/rsos.211855>

Goullier, F. (2005). *Les outils du Conseil de l'Europe en classe de langues*. Didier.

*Cadre européen commun et Portfolios*, Paris, Didier, Goullier, F. (2008). La mise en œuvre du Cadre européen commun de référence pour les langues en Europe. Une réalité différenciée dans ses finalités et dans ses modalités. *Revue internationale d'éducation de Sèvres*, 47, 55–62. <https://doi.org/10.4000/ries.367>

Gouskova, M. (2015). Optimality Theory in Phonology. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 545–566). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199677078.013.0021>

- Grabe, E., & Low, E. L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 515–546). Mouton de Gruyter.
- Graham, C. (2000). *Jazz Chants: Old and New*. Oxford University Press.
- Greenwood, P. M. (2007). Functional plasticity in cognitive aging: Review and hypothesis. *Neuropsychology*, 21(6), 657–673. <https://doi.org/10.1037/0894-4105.21.6.657>
- Grosjean, F. (2021). *Life as a Bilingual: Knowing and Using Two or More Languages*. Cambridge University Press.
- Grover, C., Jamieson, D. G., & Dobrovolsky, M. B. (n.d.). Intonation in English, French and German: Perception and production. *Language and Speech*, 33(3), 227–295. <https://doi.org/10.1177/002383098703000307>
- Guellai, B., Langus, A., & Nespors, M. (2014). Prosody in the hands of the speaker. *Frontiers in Psychology*, 5, 1–8. <https://doi.org/10.3389/fpsyg.2014.00700>
- Guberina, P. (2013). *The Verbotonal method*. Artresor Naklada, Poliklinika SUVAG.
- Guéron, J. (1974). The meter of nursery rhymes: An application of the Halle-Keyser theory of meter. *Poetics*, 3(4), 73–111. [https://doi.org/10.1016/0304-422X\(74\)90006-0](https://doi.org/10.1016/0304-422X(74)90006-0)
- Guierre, L. (1984). *Drills in English Stress Patterns*. Armand Colin-Longman.
- Guierre, L. (1987). *Règles et exercices de prononciation en anglais*. Armand Colin-Longman.
- Gut, U. (2009). Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German. Frankfurt, Germany: Peter Lang.
- Guy, G.; Horvath, B.; Vonwiller, J.; Daisley, E.; Rogers, I. (1986). An intonational change in progress in Australian English. *Language in Society*. 15: 23–52. <https://doi.org/10.1017/s0047404500011635>
- Hadding-Koch, K. & Studdert-Kennedy, M. (1964). An Experimental study of some intonation contours. *Phonetica*, 11(3-4), 175-185. <https://doi.org/10.1159/000258338>
- Hahn, L. D. (2004). Primary Stress and Intelligibility: Research to Motivate the Teaching of Suprasegmentals. *TESOL Quarterly*, 38(2), 201–223. <https://doi.org/10.2307/3588378>
- Halliday, M. (1967a). *Intonation and grammar in British English*. Mouton
- Halliday, M. (1967b). Notes on transitivity and theme in English (Parts 1–3). *Journal of Linguistics*, 3(1), 37–81; 3(2), 199–244; 4(2) (1968), 179–215.

- Hamlaoui, N., & Bengrait, N. (2016). Using Betteraccent Tutor and Praat for Learning English Intonation. *SSRN Electronic Journal: Arab World English Journal (AWE)*, 3, 99–112. <https://doi.org/10.2139/ssrn.2822981>
- Hancock, M. (2020). *Mark Hancock's 50 Tips for Teaching Pronunciation*. Cambridge University Press.
- Hansen, J. G. (2006). *Acquiring a non-native phonology: Linguistic constraints and social barriers*. Continuum.
- Harding, L. (2016). What Do Raters Need in a Pronunciation Scale? The User's View. In T. Isaacs & P. Trofimovich, (Eds.). *Assessment in second language pronunciation: Interdisciplinary perspectives* (pp. 12-34). Multilingual Matters.
- Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8(1), 34–52. <http://dx.doi.org/10.125/25228>
- Hardison, D. M. (2013). Contextualized Computer-based L2 Prosody Training: Evaluating the Effects of Discourse Context and Video Input. *CALICO Journal*, 22(2), 175–190. <https://doi.org/10.1558/cj.v22i2.175-190>
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298(5598), 1569–1579. <https://doi.org/10.1126/science.298.5598.1569>
- Hausser, R. (2004). What if Chomsky were right? *Journal of Child Language*, 31(4), 919–922. <https://doi.org/10.1017/S030500090400635X>
- Häuser, K., & Domahs, F. (2014). Functional lateralization of lexical stress representation: A systematic review of patient data. *Frontiers in Psychology*, 5, 1–4. <https://doi.org/10.3389/fpsyg.2014.00317>
- 't Hart, J., & Collier, R. (1975). Integrating different levels of intonation analysis. *Journal of Phonetics*, 3(4), 235–255. [https://doi.org/10.1016/S0095-4470\(19\)31432-9](https://doi.org/10.1016/S0095-4470(19)31432-9)
- Henderson, A, D. Frost, E. Tergujeff, A. Kautzsch, D. Murphy, A. Kirkova-Naskova, E. Waniek-Klimczak, D. Levey, U. Cunningham & L. Curnick. (2015). Pronunciation in an EFL Setting: What's going on inside & around European classrooms? *Speakout*, 52, 49-58.
- Henrichsen, L. (2019). A System for Analyzing and Evaluating Computer-Assisted Second-Language Pronunciation-Teaching Websites and Mobile Apps. In K. Graziano (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 963-968). Las Vegas, NV, United States: Association for the Advancement of Computing in Education (AACE).

Henrichsen, L. E. (2021). An Illustrated Taxonomy of Online CAPT Resources. *RELC Journal*, 52(1), 179–188. <https://doi.org/10.1177/0033688220954560>

Herry-Bénil, N. (2011). *Didactique de la phonétique anglaise*. Presses Universitaires de Rennes.

Herry, N. Hirst, D. (2002). Subjective and objective evaluation of the prosody of English spoken by French speakers: the contribution of computer assisted learning. *Proceedings of the Speech Prosody 2002 conference, Aix-en-Provence (France)*, 383–387.

Hertel, T. J., & Sunderman, G. (2009). Student Attitudes Toward Native and Non-Native Language Instructors. *Foreign Language Annals*, 42(3), 468–482. <https://doi.org/10.1111/j.1944-9720.2009.01031.x>

Hewes, G. (1973) Primate Communication and the Gestural Origin of Language. *Current Anthropology*, 14, 5-24. <https://doi.org/10.1086/201401>

Hickok G, Okada K, Barr W, Pa J, Rogalsky C, Donnelly K, Barde L, Grant A (2008). Bilateral capacity for speech sound processing in auditory comprehension: evidence from Wada procedures. *Brain and Language*. 107(3): 179–84. <https://doi.org/10.1016/j.bandl.2008.09.006>

Hickey, R. (2004). *A Sound Atlas of Irish English*. Berlin/New York: Mouton de Gruyter.

Higham, T., Compton, T., Stringer, C., Jacobi, R., Shapiro, B., Trinkaus, E., Chandler, B., Gröning, F., Collins, C., Hillson, S., O’Higgins, P., FitzGerald, C., & Fagan, M. (2011). The earliest evidence for anatomically modern humans in northwestern Europe. *Nature*, 479(7374), 521–524. <https://doi.org/10.1038/nature10484>

Hilton, H. (2009). Théories d’apprentissage et didactique des langues. *La Clé des Langues*, 12–21.

Hilton, H. (2022). *Enseigner les langues avec l’apport des sciences cognitives*. Hachette Éducation.

Hoff, E., & Shatz, M. (2007). *Blackwell handbook of language development*. Blackwell Publishing Ltd.

Holliday, A. (2006). Native-speakerism. *ELT Journal*, 60(4), 385–387. <https://doi.org/10.1093/elt/ccl030>

Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. Butler, D. B. Fry, P. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *In Honour of Daniel Jones* (pp. 73–84). Longman.

Hooper, Joan B. 1978. Constraints on schwa-deletion in American English. In Jacek Fisiak (Ed.), *Recent Developments in Historical Phonology* (pp.183-207). Norton.

Horgues, C. (2013). French Learners of L2 English: Intonation Boundaries and the Marking of Lexical Stress. *Research in Language*, 11(1), 41–56. <https://doi.org/10.2478/v10015-012-0006-8>

Hsieh, K.-T., Dong, D.-H., & Wang, L.-Y. (2013). A Preliminary Study of Applying Shadowing Technique to English Intonation Instruction. *Taiwan Journal of Linguistics*, 11(2), 43–66. [https://doi.org/10.6519/TJL.2013.11\(2\).2](https://doi.org/10.6519/TJL.2013.11(2).2)

Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) Learners: Behavioural predictors and neural substrates. *Brain and Language*, 127(3), 366–376. <https://doi.org/10.1016/j.bandl.2012.11.006>

Hübscher, I., & Prieto, P. (2019). Gestural and Prosodic Development Act as Sister Systems and Jointly Pave the Way for Children’s Sociopragmatic Development. *Frontiers in Psychology*, 10, 1259. <https://doi.org/10.3389/fpsyg.2019.01259>

Hunault, O., & Leloup, M.-H. (2022). *La première année de fonctionnement des parcours préparatoires au professorat des écoles (PPPE). Rapport à monsieur le ministre de l’Éducation nationale et de la Jeunesse madame la ministre de l’Enseignement supérieur et de la Recherche.* <https://www.education.gouv.fr/la-premiere-annee-de-fonctionnement-des-parcours-preparatoires-au-professorat-des-ecoles-pppe-343993>

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104(1), 53–69. <https://doi.org/10.1037/0033-2909.104.1.53>

Hymes, D. (1974). Ways of Speaking. *Explorations in the Ethnography of Speaking*, 1, 433- 451. <https://doi.org/10.1017/CBO9780511611810.029>

International Phonetic Association (1999). *Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet.* Cambridge University Press.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners’ L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34(3), 475–505. <https://doi.org/10.1017/S0272263112000150>

Isaacs, T., & Trofimovich, P. (2022). Reactions to second language speech. Influences of discrete speech characteristics, rater experience, and speaker first language background. In J. Levis, T. Derwing & M. Munro (Eds) *The evolution or pronunciation teaching and research. 25 years of intelligibility, comprehensibility, and accentedness* (pp. 125-151). John Benjamins.



- Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, 97(1), 553–562. <https://doi.org/10.1121/1.412280>
- Iversen, J. R., Patel, A. D., & Ohgushi, K. (2006). *How the Mother Tongue Influences the Musical Ear*. 7. [https://acoustics.org/pressroom/httpdocs/152nd/iversen\\_patel\\_ohgushi.html](https://acoustics.org/pressroom/httpdocs/152nd/iversen_patel_ohgushi.html)
- Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis; the distinctive features and their cues*. MIT Press.
- James, E. F. (1976). The acquisition of prosodic features of speech using a speech visualizer. *IRAL - International Review of Applied Linguistics in Language Teaching*, 14(3). <https://doi.org/10.1515/iral.1976.14.3.227>
- Javet, B. (Ed.). (2021). *Repères et références statistiques : Enseignements, formation, recherche : [RERS 2021]*. Ministère de l'Éducation Nationale et de la Jeunesse : Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. Retrieved 10 May, 2023 from <https://www.education.gouv.fr/reperes-et-references-statistiques-2021-308228>
- Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83 - 103. <https://doi.org/10.1093/applin/23.1.83>
- Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, 24, 109–125. <https://doi.org/10.1017/S0267190504000054>
- Jenkins, J. (2009). Who Speaks English Today? In *World Englishes. A Resource Book for Students* (pp. 15–24). Routledge.
- Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching*, 44(3), 281–315. <https://doi.org/10.1017/S0261444811000115>
- Jenkins, J., Baker, W., & Dewey, M. (Eds.) (2018). *The Routledge handbook of English as a Lingua Franca*. Routledge/Taylor & Francis Group.
- Jenkins, R. (1961) Perception of Pitch, Timbre and Loudness, *The Journal of the Acoustical Society of America*, 33(11), 1550-1557.

- Jeong, H., Lindemann, S., & Forsberg, J. (2021). *English phonology in a globalized world: Challenging native speakerism through listener training in universities in Sweden and the US* [Preprint]. SocArXiv. <https://doi.org/10.31235/osf.io/4sjtz>
- Jespersen, O. (1933). *Notes on metre*. Reprinted in H. Gross (Ed) (1979). *The structure of verse* (pp. 105–128). Ecco.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). Guilford Press.
- John, R. Robins, W., Pervin, L.A. (Eds.) (2008), *Handbook of personality: Theory and research*. Guilford Press.
- Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The Edge Factor in Early Word Segmentation: Utterance-Level Prosody Enables Word Form Extraction by 6-Month-Olds. *PLoS ONE*, 9(1), e83546. <https://doi.org/10.1371/journal.pone.0083546>
- Jolley, C. (2014). *The Effect of Computer-Based Pronunciation Readings on ESL Learners' Perception and Production of Prosodic Features in a Short-Term ESP Course* [MA thesis]. Brigham Young University.
- Jones, D. (1918). *An Outline of English Phonetics*. Heffer and Sons Ltd.
- Jordan, E. (2011). Regional International Englishes – The Future of English as a Lingua Franca? *The International Journal - Language Society and Culture*, 33, 30–36.
- Jun, S & Fougeron, C. (1995). The accentual phrase and the prosodic structure of French. *13th International Congress of Phonetic Sciences (ICPhS 12)*, vol. 2, 722–725.
- Jun, Sun-Ah & Fougeron, C. (2000). A phonological model of French intonation. In Antonis Botinis (Ed.), *Intonation: Analysis, modeling and technology* (pp. 209–242). Kluwer.
- Jung, Y., Kim, Y., & Murphy, J. (2017). The role of task repetition in learning word-stress patterns through auditory priming tasks. *Studies in Second Language Acquisition*, 39(2), 319–346. <https://doi.org/10.1017/S0272263117000031>
- Jungers, M. K., Hupp, J. M., & Dickerson, S. D. (2016). Language Priming by Music and Speech. *Music Perception*, 34(1), 33–39. <https://doi.org/10.1525/mp.2016.34.1.33>
- Kachru, B. (1985). Standards, codification and sociolinguistic realism: English language in the outer circle. In R. Quirk and H. Widowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-36). Cambridge University Press.

- Kamiyama, T., & Shinohara, S. (2010). Role of prosody in segmental perception in L2. *Sophia University Working Papers in Phonetics 2010*, 10–18. <https://doi.org/10.1.1.608.4962>
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315. <https://doi.org/10.1016/j.system.2010.01.005>
- Kang, O., & Ginther, A. (Eds.). (2018). *Assessment in second language pronunciation*. Routledge.
- Kang, O., Vo, S. C. T., & Moran, M. K. (2016). Perceptual Judgments of Accented Speech by Listeners from Different First Language Backgrounds. *TESL-EJ*, 20(1), 1–24.
- Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical Approaches to Measuring the Intelligibility of Different Varieties of English in Predicting Listener Comprehension: Measuring Intelligibility in Varieties of English. *Language Learning*, 68(1), 115–146. <https://doi.org/10.1111/lang.12270>
- Kang, O., Johnson, D. O., & Kermad, A. (2021). *Second Language Prosody and Computer Modeling* (First edition). Routledge. <https://doi.org/10.4324/9781003022695>
- Karlsson, F., & Hartelius, L. (2021). On the Primary Influences of Age on Articulation and Phonation in Maximum Performance Tasks. *Languages*, 6(4), 174. <https://doi.org/10.3390/languages6040174>
- Kaulfers, W. (1929). An Inductive Method of Teaching Pronunciation. *The Modern Language Journal*, 13(8), 610–619.
- Kember, H., Grohe, A.-K., Zahner, K., Braun, B., Weber, A., & Cutler, A. (2017). Similar Prosodic Structure Perceived Differently in German and English. *Interspeech 2017*, 1388–1392. <https://doi.org/10.21437/Interspeech.2017-544>
- Kerswill, P. (2006). RP, Standard English and the standard/non-standard relationship. In Britain, D. (Ed.), *Language in the British Isles*, (pp. 34–51). Cambridge University Press.
- Kennedy, S., & Trofimovich, P. (2019). Comprehensibility: A Useful Tool to Explore Listener Understanding. *The Canadian Modern Language Review*, 75(4), 275–284. <https://doi.org/10.3138/cmlr.2019-0280>
- Kenworthy, J. (1987). *Teaching English Pronunciation*. Longman.
- Kissau, S. (2006). Gender differences in second language motivation: An investigation of micro- and macro-level influences. *Canadian Journal of Applied Linguistics*, 9(1), 73–96.
- Kjellin, O. (1999). Accent Addition: Prosody and Perception Facilitates Second Language Learning. In O. Fujimura, B. D. Joseph, & B. Palek (Eds.), *Proceedings of LP'98 (Linguistics and Phonetics Conference)*

at Ohio State University, Columbus, Ohio, September 1998 (Vol. 2) Prague: The Karolinum Press, 373-398.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038–1054. <https://doi.org/10.1121/1.1923349>

Kolly, M.-J., Boula de Mareüil, P., Leemann, A., & Dellwo, V. (2017). Listeners use temporal information to identify French- and English-accented speech. *Speech Communication*, *86*, 121–134. <https://doi.org/10.1016/j.specom.2016.11.006>

Komatsu, M., Mori, K., Arai, T., Aoyagi, M., & Murahara, Y. (2002). Human language identification with reduced segmental information. *Acoustical Science and Technology*, *23*(3), 143–153. <https://doi.org/10.1250/ast.23.143>

Kostromitina, M., & Kang, O. (2021). The Effects of ESL Immersion and Proficiency on Learners' Pronunciation Development. *Frontiers in Communication*, *6*, 636122. <https://doi.org/10.3389/fcomm.2021.636122>

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The Evolution of Rhythm Processing. *Trends in Cognitive Sciences*, *22*(10), 896–910. <https://doi.org/10.1016/j.tics.2018.08.002>

Krashen, S. D. (1985). *Second language acquisition and second language learning*. Pergamon.

Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831-843. <https://doi.org/10.1038/nrn1533>

Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, *67*(5), 713–727. <https://doi.org/10.1016/j.neuron.2010.08.038>

Kurt, B. G., Medlin, J., & Tessarolo, A. (2014). The Perception of Prosodically Ambiguous Intonation Patterns by L2 English Learners and the Effects of Instruction. *Concordia Papers in Working Linguistics*, *5*, 353–372.

Lacheret-Dujour, A., Beaugendre, F. (1999). *La prosodie du français*. CNRS.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge University Press.

Lado, R. (1957). *Linguistics across cultures*. University of Michigan Press.

Lakoff, R. (1973). Language and woman's place. *Language in Society* 2:45–80.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.

Lapaire, J-R. (2006). *La grammaire anglaise en mouvement*. Hachette.

Lapaire, J-R. (2010). Postures, manipulations, déambulations : comprendre la grammaire anglaise autrement. *La nouvelle revue de l'adaptation et de la scolarisation*, 49, 1-14.

Lapaire, J.-R. (2014a). À corps perdu ou le mystère de la désincarnation des langues. *E-CRINI, la revue électronique du Centre de Recherche sur les Identités*, 6, 1–16.

Lapaire, J.-R. (2014b). Grammaire de l'oral et engagement du corps apprenant. In C. Martinot & A. Pégaz Paquet (Eds.), *Innovations didactiques en français langue étrangère* (pp. 25–37). CRL.

Larsson, M., Richter, J., & Ravignani, A. (2019). Bipedal steps in the development of rhythmic behavior in humans. *Music & Science*, 2, 1-14. <https://doi.org/10.1177/2059204319892617>

Lasagabaster, D., & Pagèze, J. (2018). Teacher Development for Teaching and Learning in English in a French Higher Education Context. In J. Valcke, A. C. Murphy, & F. Costa (Eds.), *Critical Issues in English-Medium Instruction at University*. Facoltà di Scienze Linguistiche e Letterature straniere Università Cattolica del Sacro Cuore, pp. 289–310).

Lascotte, D. K., & Tarone, E. (2022). Channeling Voices to Improve L2 English Intelligibility. *The Modern Language Journal*, 106(4), 744–763. <https://doi.org/10.1111/modl.12812>

Lau, J. C. Y., Fyshe, A., & Waxman, S. R. (2022). Rhythm may be key to linking language and cognition in young infants: evidence from machine learning. *Frontiers in Psychology*, 13, 894405. <https://doi.org/10.3389/fpsyg.2022.894405>

Leather, J. (2002). Modeling the acquisition of speech in a “multilingual” society: An ecological approach. In Kramsch, C. (Ed.) *Language acquisition and language socialization: ecological perspectives* (pp. 325). Continuum.

Lecanuet, J.-P., & Schaal, B. (2002). Sensory performances in the human foetus: A brief summary of research. *Intellectica. Revue de l'Association Pour La Recherche Cognitive*, 34(1), 29–56. <https://doi.org/10.3406/intel.2002.1072>

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366. <https://doi.org/10.1093/applin/amu040>

- Legendre, J. (2003). *Rapport d'information fait au nom de la commission des Affaires culturelles sur l'enseignement des langues étrangères en France*. Le Sénat. <https://www.senat.fr/rap/r03-063/r03-0631.pdf>
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5(3), 253–263. [https://doi.org/10.1016/S0095-4470\(19\)31139-8](https://doi.org/10.1016/S0095-4470(19)31139-8)
- Lehiste, I., & Peterson, G. E. (1959). Vowel Amplitude and Phonemic Stress in American English. *The Journal of the Acoustical Society of America*, 31(4), 428–435. <https://doi.org/10.1121/1.1907729>
- Lehmann, W.P. (1956). *Proto-Indo-European Phonology*. University of Texas Press and Linguistic Society of America.
- Lehmann, W. P. (1956). *The development of Germanic verse form*. University of Texas Press.
- Lenneberg, E.H. (1967). *Biological Foundations of Language*. Wiley and Sons.
- Leonard, L. B. (1973). The Role of Intonation in the Recall of Various Linguistic Stimuli. *Language and Speech*, 16(4), 327–335. <https://doi.org/10.1177/002383097301600403>
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75. <https://doi.org/10.1017/S0140525X99001776>
- Levis, J. M. (2005). Changing Contexts and Shifting Paradigms in Pronunciation Teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.1075/jslp.20050.lev>
- Levis, J (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge University Press.
- Levis, J. (2020). Revisiting the Intelligibility and Nateness Principles. *Journal of Second Language Pronunciation*, 6(3), 310–328. <https://doi.org/10.1075/jslp.20050.lev>
- Levis, J. (2022). Evolution of L2 pronunciation research and teaching: 25 years of intelligibility, comprehensibility, and accentedness. In J. Levis, T. Derwing & M. Munro (Eds) *The evolution or pronunciation teaching and research. 25 years of intelligibility, comprehensibility, and accentedness* (pp. 1-5). John Benjamins.

- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
- Lieberman, P. (1960). Some Acoustic Cues of Word Stress in American English. *The Journal of the Acoustical Society of America*, 32, 451–454. <https://doi.org/10.1121/1.1908095>
- Lilly, R. Viel, M. (1998a) *Initiation raisonné à la phonétique de l'anglais*. Hachette Supérieur.
- Lilly, R., Viel, M. (1998b) *La prononciation de l'anglais : Règles phonologiques et exercices de transcription*. Hachette Supérieur.
- Lima, E. de F. (2015). *Development and evaluation of online pronunciation instruction for international teaching assistants' comprehensibility* [PhD thesis, Iowa State University, Digital Repository]. <https://doi.org/10.31274/etd-180810-4109>
- Lin, P. M. S. (2012). Sound Evidence: The Missing Piece of the Jigsaw in Formulaic Language Research. *Applied Linguistics*, 33(3), 342–347. <https://doi.org/10.1093/applin/ams017>
- Lippi-Green, R. (1994). Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society*, 23, 163–198. <https://doi.org/10.1017/S0047404500017826>
- Liu, D., & Reed, M. (2021). Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study. *Frontiers in Communication*, 6, 627316. <https://doi.org/10.3389/fcomm.2021.627316>
- Llorca, R. (1998). *Les Ritmimots, exercices de groupes avec la voix et le geste sur les rythmes du français parlé*. Bayerischer Rundfunk.
- Llurda, E. (2004). Non-native-speaker teachers and English as an International Language. *International Journal of Applied Linguistics*, 14(3), 314–323. <https://doi.org/10.1111/j.1473-4192.2004.00068.x>
- Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 3(2), 157–189. [https://doi.org/10.1016/0030-5073\(68\)90004-4](https://doi.org/10.1016/0030-5073(68)90004-4)
- Locke, E. A., & Latham, G. P. (2006). New Directions in Goal-Setting Theory. *Current Directions in Psychological Science*, 15(5), 265–268. <https://doi.org/10.1111/j.1467-8721.2006.00449.x>
- Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & Cognition*, 42, 41–52. <https://doi.org/10.3758/s13421-013-0342-5>
- Manès-Bonnisseau, C. (2019). *Guide pour l'enseignement des langues vivantes Oser les langues vivantes étrangères à l'école*. Ministère de l'Éducation nationale et de la Jeunesse. Retrieved 10 May,

2023 from <https://eduscol.education.fr/159/guide-pour-l-enseignement-des-langues-vivantes-etrangeres>

Maley, A. (2000). The language teacher's voice. Macmillan Heinemann English Language Teaching

McAndrews, M. (2019). Short periods of instruction improve learners' phonological categories for L2 suprasegmental features. *System*, 82, 151–160. <https://doi.org/10.1016/j.system.2019.04.007>

McAndrews, M. M., & Thomson, R. I. (2017). Establishing an empirical basis for priorities in pronunciation teaching. *Journal of Second Language Pronunciation*, 3(2), 267–287. <https://doi.org/10.1075/jslp.3.2.05mca>

McArthur, T. (1987). The English languages? *English Today*, 3(3), 9–13. <https://doi.org/10.1017/S0266078400013511>

McCrocklin, S., & Link, S. (2016). Accent, Identity, and a Fear of Loss? ESL Students' Perspectives. *The Canadian Modern Language Review*, 72(1), 122–148. <https://doi.org/10.3138/cmlr.2582>

MacDougall, H. G., & Moore, S. T. (2005). Marching to the beat of the same drummer: The spontaneous tempo of human locomotion. *Journal of Applied Physiology*, 99, 1164–1173. <https://10.1152/jappphysiol.00138.2005>

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748. <https://doi.org/10.1038/264746a0>

Mckenzie, R. T. (1905). The facial expression of violent effort, breathless. *Journal of Anatomy and Physiology*, 40, 51–55.

Mcmahon, A. (2007). Sounds, brain, and evolution: or, why phonology is plural. In: Pennington, M.C. (eds) *Phonology in Context*. Palgrave Advances in Linguistics. Palgrave Macmillan, London. [https://doi.org/10.1057/9780230625396\\_7](https://doi.org/10.1057/9780230625396_7)

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21(4), 499–511. <https://doi.org/10.1017/S0140525X98001265>

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press.

MacWhinney, B. (1987). The Competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Lawrence Erlbaum Associates, Inc.



- MacWhinney, B. (1992). Transfer and Competition in Second Language Learning. *Advances in Psychology*, 83, 371–390. Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)61506-X](https://doi.org/10.1016/S0166-4115(08)61506-X)
- MacWhinney, B. (2005). A Unified Model of Language Acquisition. In J.F. Kroll & A.M.B. De Groot (Eds) *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford University Press.
- MacWhinney, B. (2018). A unified model of first and second language learning. In M. Hickmann, E. Veneziano, & H. Jisa (Eds.), *Trends in Language Acquisition Research* (Vol. 22, pp. 287–312). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.22.15mac>
- MacWhinney, Brian (2019). Neuroemergentism: Levels and constraints. *Journal of Neurolinguistics*. 49, 232–234. <https://doi.org/10.1016/j.jneuroling.2018.04.002>
- MacWhinney, B., & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
- Maca-Meyer, N., González, A. M., Larruga, J. M., Flores, C., & Cabrera, V. M. (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics*, 2(13), 1–8. <https://doi.org/10.1186/1471-2156-2-13>
- Magen, H. S. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, 26(4), 381–400. <https://doi.org/10.1006/jpho.1998.0081>
- Mahn, H. and John-Steiner, V. (2012). Vygotsky and Sociocultural Approaches to Teaching and Learning. In I. Weiner, W.M. Reynolds and G.E. Miller (Eds.) *Handbook of Psychology, Second Edition* (pp. 125-151). <https://doi.org/10.1002/9781118133880.hop207006>
- Mahncke H.W., Bronstone A., Merzenich M.M. (2006). Brain plasticity and functional losses in the aged: scientific bases for a novel intervention. *Progress in Brain Research*, 157, 81-109. [https://doi:10.1016/S0079-6123\(06\)57006-2](https://doi:10.1016/S0079-6123(06)57006-2)
- Martin, J. G. (1968). Temporal word spacing and the perception of ordinary, anomalous, and scrambled strings. *Journal of Verbal Learning and Verbal Behavior*, 7(1), 154–157. [https://doi.org/10.1016/S0022-5371\(68\)80181-1](https://doi.org/10.1016/S0022-5371(68)80181-1)
- Martin, J. G. (1979). Rhythmic and segmental perception are not independent. *Journal of the Acoustical Society of America*, 65(5), 1286–1297.
- Martin, P. (2005). WinPitch LTL, un logiciel multimédia d'enseignement de la prosodie. *Alsic*, 8(2). <https://doi.org/10.4000/alsic.332>

Martin, P. (2019). Enseignement de l'intonation en FLE aujourd'hui. *Recherches en didactique des langues et des cultures*, 16(1). <https://doi.org/10.4000/rdlc.4431>

Martinez-Flor, A. & Usó-Juan, E. (2006). Towards acquiring communicative competence through listening. In E. Usó-Juan and A. Martínez-Flo (Eds.) *Current Trends in the Development and Teaching of the four Language Skills* (pp. 29-40). De Gruyter Mouton,

Meltzer, R. H., Martin, J. G., Mills, C. B., Imhoff, D. L., & Zohar, D. (1976). Reaction time to temporally-displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance*, 2(2), 277–290. <https://doi.org/10.1037/0096-1523.2.2.277>

Mennen, I., & de Leeuw, E. (2014). Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition*, 36(2), 183–194. <https://doi.org/10.1017/S0272263114000138>

Mercer, S. (2018). Psychology for language learning: Spare a thought for the teacher. *Language Teaching*, 51(4), 504–525. <https://doi.org/10.1017/S0261444817000258>

Merriënboer, J. J. G. van, & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions Authors(s): Jeroen J. G. van Merriënboer and John Sweller. *Educational Psychology Review*, 17(2), 147–177. <https://doi.org/10.1007/s10648-005-3951-0>

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P.A.A, & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Research* 1194, 81-89. <http://doi.org/10.1016/j.brainres.2007.11.042>

Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language and Speech*, 21(3), 221–241. <https://doi.org/10.1177/002383097802100302>

Ministère de l'Éducation Nationale, de la Jeunesse et des Sports. (2003). Organisation des études conduisant au DUT de la spécialité informatique : *Bulletin officiel N°20 du 15 mai 2003, annexe*. Retrieved 10 May, 2023 from <https://www.education.gouv.fr/bo/2003/20/sup.htm>

Ministère de l'Éducation Nationale, de la Jeunesse et des Sports. (2006). *Le socle commun des connaissances et des compétences : décret du 11 juillet 2006*. Ministère de l'Éducation Nationale, 2006. Retrieved 10 May, 2023 from <http://media.education.gouv.fr/file/51/3/3513.pdf>

Ministère de l'Éducation Nationale, de la Jeunesse et des Sports. (2007). Mise en œuvre du socle commun de connaissances et de compétences : Programmes d'enseignement de l'école primaire :

arrêté du 4-4-2007. *Bulletin officiel*, 12 avril 2007, hors-série (5). Retrieved 10 May, 2023 from <http://www.education.gouv.fr/bo/2007/hs5/default.htm>

Ministère de l'Éducation Nationale, de la Jeunesse et des Sports. (2020). Programmes d'enseignement Cycle des apprentissages fondamentaux (cycle 2), cycle de consolidation (cycle 3) et cycle des approfondissements (cycle 4) : modification. *Bulletin Officiel 31 du 30 juillet, 2020*. Retrieved 10 May, 2023 from [https://www.education.gouv.fr/pid285/bulletin\\_officiel.html?pid\\_bo=39771](https://www.education.gouv.fr/pid285/bulletin_officiel.html?pid_bo=39771)

Ministère de l'Éducation Nationale, de la Jeunesse et des Sports. (2022). Enseignement de l'anglais et des langues vivantes étrangères tout au long de la scolarité obligatoire Mesures pour améliorer les apprentissages des élèves. Circulaire du 12-12-2022. Retrieved 10 May, 2023 from <https://www.education.gouv.fr/bo/22/Hebdo47/MENE2234752C.htm>

Miras, G. (2014). *Approche plurielle des liens musique-parole pour la didactique de la prononciation du français comme langue étrangère/seconde* [PhD thesis]. Sorbonne Nouvelle, Paris 3.

Miras, G. (2019). De la correction à la médiation : La doxa terminologique en didactique de la prononciation du français comme langue étrangère. *Recherches en didactique des langues et des cultures*, 16(1). <https://doi.org/10.4000/rdlc.4298>

Miras, G., Wilson, A., & Dupouy, M. (2022). Utopie et dystopie hexagonales sur l'accent et la prononciation (du natif) en langues étrangères. *Humanités, Didactiques, Recherches*, 2, 117–133.

Missaglia, F. (1999). Contrastive prosody in SLA – an empirical study with adult Italian learners of German. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th Congress of Phonetic Sciences, 1* (Vol. 1, pp. 551–554). The Regents of the University of California. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/index.html>

Mithen, S. (2005). *The singing Neanderthals: The origins of music, language, mind, and body*. Weidenfeld & Nicolson.

Mithen, S. (2011). The significance of stones and bones: understanding the biology and evolution of rhythm requires attention to the archaeological and fossil record. In P. Rebuschat, M. Rohmeier, J.A. Hawkins, I Cross (Eds.). *Language and Music as Cognitive Systems*. Oxford University Press.

Molaie, S. (2018). Apprentissage de l'anglais en contexte universitaire : motivation, créativité et rétention. (PhD thesis), Université Grenoble Alpes.

Mompeán, J. A. (2014). Stress shift in English: The case of teen numbers. In R. Monroy-Casas & I. Arboleda-Guirao (Eds.), *Readings in English Phonetics and Phonology* (pp. 149-164). IULMA.

Moosmüller, S. (1989). Phonological variation in parliamentary discourse. In R Wodak (Ed.), *Language, power and ideology: Studies in political discourse* (pp.165–180). John Benjamins.

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Psychology Press.

Moustapha-Sabeur, M. (2016). Faire parler l'apprenant en classe de FLE : Le rôle et l'influence des indices prosodiques. *Recherches en didactique des langues et des cultures*, 13(2). <https://doi.org/10.4000/rdlc.826>

Moyer, A. (2014). Exceptional Outcomes in L2 Phonology: The Critical Factors of Learner Engagement and Self-Regulation. *Applied Linguistics*, 35(4), 418–440. <https://doi.org/10.1093/applin/amu012>

Mugler, E. M., Patton, J. L., Flint, R. D., Wright, Z. A., Schuele, S. U., Rosenow, J., Shih, J. J., Krusienski, D. J., & Slutzky, M. W. (2014). Direct classification of all American English phonemes using signals from functional speech motor cortex. *Journal of Neural Engineering*, 11(3), 035015. <https://doi.org/10.1088/1741-2560/11/3/035015>

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>

Munro, M. J., & Derwing, T. M. (1998). The Effects of Speaking Rate on Listener Evaluations of Native and Foreign-Accented Speech. *Language Learning*, 48(2), 159–182. <https://doi.org/10.1111/1467-9922.00038>

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520–531. <https://doi.org/10.1016/j.system.2006.09.004>

Munro, M. J., & Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching*, 44(3), 316–327. <https://doi.org/10.1017/S0261444811000103>

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, 44(1), 1–35. <https://doi.org/10.1017/S0261444810000327>

Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, 42, 258–269. <https://doi.org/10.1016/j.system.2013.12.007>

Murphy, J. M. (2017). Teacher training in the teaching of pronunciation. In O. Kang, R. Thomson, & J. M. Murphy (Eds.), *The Routledge Handbook of Second Language Pronunciation* (pp. 298–319). Routledge/Taylor & Francis Group.

- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756–766. [https://doi.org/0096-1523/98/\\$3.00](https://doi.org/0096-1523/98/$3.00)
- Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behaviour & Development*, 21(4), 779–784.
- Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (Ed.), *Logical issues in language acquisition*, (pp. 157-175). Foris.
- Nida, E. (1957). *Learning a foreign language*. Friendship.
- Niebuhr, O. (2021). Computer-assisted prosody training: Improving public speakers' vocal charisma with the Web-Pitcher. *Revista Da Abralin*, 1–29. <https://doi.org/10.25189/rabralin.v20i1.1809>
- Nocaudie, O., Alazard-Guiou, C., & Billières, M. (2019). Oral d'aujourd'hui, oralité de demain : Et la phonétique corrective dans tout cela ? *Recherches en didactique des langues et des cultures*, 16(1). <https://doi.org/10.4000/rdlc.4615>
- Nooteboom, S. (1997). The Prosody of Speech: Melody and Rhythm, in W. Hardcastle & J. Laver, (Eds.), *The Handbook of Phonetic Sciences*, (pp. 640-673). Blackwell.
- North, B. (2020). The CEFR renewed: Inspiring the future of language education. *Italiano LinguaDue*, 1, 548–560.
- Nunan, D. (2012). *Learner-Centered English Language Education: The Selected Works of David Nunan*. Routledge. <https://doi.org/10.4324/9780203096888>
- Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112(1), 181–186. <https://doi.org/10.1016/j.cognition.2009.04.001>
- O'Brien, M. G. & Levis, J. M. (2017). Pronunciation and technology. In M. O'Brien & J. Levis (Eds). *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, Calgary, AB, August 2016 (pp. 1-9). Ames, IA: Iowa State University.
- OECD (2018), "Embodied learning", in *Teachers as Designers of Learning Environments: The Importance of Innovative Pedagogies*, OECD Publishing, <https://doi.org/10.1787/9789264085374-11-en>
- O'Neil, D. (2019). The Middle English Creolization Hypothesis: Persistence, Implications, and Language Ideology. *Studia Anglica Posnaniensia*, 54(1), 113–132. <https://doi.org/10.2478/stap-2019-0006>

- Ohala, J. J., & Gilbert, J. B. (1981). Listeners' ability to identify languages by their prosody. In M. Rossi & P. R. Léon (Eds.), *Problèmes de prosodie : Hommages à Georges Faure. Vol. 2, Expérimentations, modèles et fonctions / édité par Pierre Léon et Mario Rossi*. Didier.
- Ordin, M., Polyanskaya, L., & Ulbrich, C. (2011). Acquisition of timing patterns in second language. *Interspeech 2011*, 1129–1132. <https://doi.org/10.21437/Interspeech.2011-336>
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Heinle & Heinle.
- Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5(3), 261–283. <https://doi.org/10.1007/BF01067377>
- Pakkenberg, B. (2003). Aging and the human neocortex. *Experimental Gerontology*, 38(1–2), 95–99. [https://doi.org/10.1016/S0531-5565\(02\)00151-1](https://doi.org/10.1016/S0531-5565(02)00151-1)
- Palmer, C., & Kelly, M. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language*, 31(4), 525–542. [https://doi.org/10.1016/0749-596X\(92\)90027-U](https://doi.org/10.1016/0749-596X(92)90027-U)
- Paradis, J. (2007). Second language acquisition in childhood. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 387–405). Blackwell Publishing. <https://doi.org/10.1002/9780470757833.ch19>
- Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1(3), 213–237. <https://doi.org/10.1075/lab.1.3.01par>
- Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Percept.* 24, 99–104. <https://doi.org/10.1525/mp.2006.24.1.99>
- Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford University Press.
- Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00142>
- Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, 119(5), 3034–3047. <https://doi.org/10.1121/1.2179657>
- Pater, J. (1993). Theory and methodology in the study of metrical parameter (re)setting. *McGill Working Papers in Linguistics*, 9, 211–243.

- Pater, J. (1997). Metrical parameter in setting in second language acquisition. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on Phonological Acquisition*, (pps. 235-261). John Benjamins.
- Patkowski, Mark. S. (1990). Age and Accent in a Second Language: A Reply to James Emil Flege. *Applied Linguistics*, 11(1), 73–89. <https://doi.org/10.1093/applin/11.1.73>
- Paivio, A. (1963). Learning of adjective-noun paired associates as a function of adjective-noun word order and noun abstractness. *Canadian Journal of Psychology / Revue canadienne de psychologie*, 17(4), 370–379. <https://doi.org/10.1037/h0083277>
- Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford University Press.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, 45(3), 255–287. <https://doi.org/10.1037/h0084295>
- Paivio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology*, 80(2, Pt.1), 279–285. <https://doi.org/10.1037/h0027273>
- Pavlov, P., Svendsen, J. I., & Indrelid, S. (2001). Human presence in the European Arctic nearly 40,000 years ago. *Nature*, 413(6851), 64–67. <https://doi.org/10.1038/35092552>
- Patey, A. (2010). *Activités théâtrales en classe de langue*. CLE International.
- Pedrazzini, L. (2016). The Principle of ‘Correct Pronunciation’: Teaching English as a Foreign Language in the Early Twentieth Century. *Language & History*, 59(1), 63–72. <https://doi.org/10.1080/17597536.2016.1176292>
- Pellegrino, E. (2012). The Perception of Foreign Accented Speech. Segmental and Suprasegmental features affecting the degree of foreign accent in L2 Italian. In H. Mello, M. Pettorino, & T. Raso (Eds.), *Proceedings of the 8th GSCP Conference* (pp. 261–267). Firenze University Press.
- Pellegrino, E., Schwab, S., & Dellwo, V. (2021). Native listeners rely on rhythmic cues when deciding on the nativeness of speech. *The Journal of the Acoustical Society of America*, 150(4), 2836–2853. <https://doi.org/10.1121/10.0006537>
- Pennington, M. C. (1989). Teaching Pronunciation from the Top Down. *RELC Journal*, 20(1), 20–38. <https://doi.org/10.1177/003368828902000103>
- Pennington, M. C. (Ed.). (2007). *Phonology in context*. Palgrave Macmillan.
- Pennington MC (2019). ‘Top-down’ pronunciation teaching revisited. *RELC Journal* 50(3): 371-385.

- Pennington, M. C. (2021). Teaching Pronunciation: The State of the Art 2021. *RELC Journal*, 52(1), 3–21. <https://doi.org/10.1177/00336882211002283>
- Pennington, M. C., & Ellis, N. C. (2000). Cantonese Speakers' Memory for English Sentences with Prosodic Cues. *The Modern Language Journal*, 84(3), 372–389. <https://doi.org/10.1111/0026-7902.00075>
- Pennington, M. C., & Rogerson-Revell, P. (2019). *English Pronunciation Teaching and Research: Contemporary Perspectives*. Palgrave Macmillan UK. <https://doi.org/10.1057/978-1-137-47677-7>
- Pépiot, E. (2015). Voice, speech and gender: Male-female acoustic differences and cross-language variation in English and French speakers. *Corela*, HS-16, 1-14. <https://doi.org/10.4000/corela.3783>
- Peppé, S. (2013). Profiling Elements of Prosody in Speech-Communication (PEPS-C). In Volkmar, F. (Ed.) *Encyclopedia of Autism Spectrum Disorders*. Springer.
- Pettorino, M., Meo, A. D., & Vitale, M. (2012). Transplanting credibility into a foreign voice: An experiment on synthesized L2 Italian. In H. Mello, M. Pettorino, & T. Raso (Eds.), *Proceedings of the 7th GSCP International Conference* (pp. 281–284). Firenze University Press.
- Phillips, S., Aguilar Perez, A., Alt, H., & Darcy, I. (2022). Pause for thought (groups): non-native pausing behavior and ease of processing of L2 speech. In J. Levis & A. Guskaroska (Eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, <https://doi.org/10.31274/psllt.13355>
- Picavet, F., Aubergé, V., & Rossato, S. (2013). Can a guided rhythmic approach contribute to the oral performance of learners of L2 English? A case study. In M. G. Busa & A. Stella (Eds.), *Methodological Perspectives on Second Language Prosody Papers from ML2P 2012* (p. 7). CLEUP Sc.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. [PhD thesis], Massachusetts Institute of Technology.
- Pierrehumbert, J. & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–323). MIT Press.
- Pike, K. L. (1945). *The intonation of American English*. University of Michigan Press.
- Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, 143(1), 195–204. <https://doi.org/10.1037/a0032246>



Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6), 322–334. <https://doi.org/10.1038/s41583-020-0304-4>

Polivanov E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague*. (4), 79–96.

Polyanskaya, L., Samuel, A. G., & Ordin, M. (2019). Speech Rhythm Convergence as a Social Coalition Signal. *Evolutionary Psychology*, 17(3), 147470491987933. <https://doi.org/10.1177/1474704919879335>

Pommée, T. (2021). *Les mesures d'intelligibilité : État de l'art, considérations pratiques pour l'applicabilité clinique et explorations acoustiques* [PhD thesis]. Retrieved 10 May, 2023 from [http://www.afcp-parole.org/wp-content/uploads/2021/12/POMMEE\\_Manuscrit\\_these\\_FINAL\\_CINES\\_red.pdf](http://www.afcp-parole.org/wp-content/uploads/2021/12/POMMEE_Manuscrit_these_FINAL_CINES_red.pdf)

Pommée, T., Balaguer, M., Mauclair, J., Piquier, J., & Woisard, V. (2022). Intelligibility and comprehensibility: A Delphi consensus study. *International Journal of Language & Communication Disorders*, 57(1), 21–41. <https://doi.org/10.1111/1460-6984.12672>

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816–847. <https://doi.org/10.1016/j.neuroimage.2012.04.062>

Prince, A., & Smolensky, P. (1993). *Constraint Interaction in Generative Grammar* (p. 262) [RuCCS Technical Report].

Privas-Bréauté, V. (2018). Le jeu et les techniques dramatiques en classe de langue : Un levier pour la co-construction des compétences professionnelles. *Mélanges CRAPEL (n° Varia)*, 38(2), 77–94.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576–582. <https://doi.org/10.1038/nrn1706>

Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, 30(2), 271–287. <https://doi.org/10.1111/j.1467-1770.1980.tb00319.x>

Puren, C. (1994). *La didactique des langues étrangères à la croisée des méthodes : Essai sur l'éclecticisme*. Didier.

- Quesada Vázquez, L. (2021). Teaching English Pronunciation Online during the COVID-19 Crisis Outbreak. *7th International Conference on Higher Education Advances (HEAd'21)*. Editorial Universitat Politècnica de València, 351-358. <https://doi.org/10.4995/HEAd21.2021.12906>
- Quesada Vázquez, L. (2022). Synchronous computer-mediated communication in English pronunciation teaching: A case study of Rovira i Virgili University. *Revista de Estilos de Aprendizaje / Journal of Learning Styles*, 20(10), 123–134.
- Quirk, R., & Widdowson, H. G. (Eds.). (1985). *English in the World: Teaching and Learning the Language and Literatures*. Cambridge University Press for the British Council.
- Ramus, F., Nespors, M., & Mehler, J. (2000). Cues of linguistic rhythm in the speech signal. *Cognition*, 75(1), AD3–AD30. [https://doi.org/10.1016/S0010-0277\(00\)00101-3](https://doi.org/10.1016/S0010-0277(00)00101-3)
- Randall, M. (2007). *Memory, psychology and second language learning*. John Benjamins.
- Rao, R. (2009). Deaccenting in spontaneous speech in Barcelona Spanish. *Studies in Hispanic and Lusophone Linguistics*, 2, 31–75. <https://doi.org/10.1515/shll-2009-1035>
- Rapin, L., & Ménard, L. (2019). The Multimodal Perception of Contrastive Focus in French: A Developmental Study. *Frontiers in Communication*, 3, 60. <https://doi.org/10.3389/fcomm.2018.00060>
- Rasier, L., & Hilgsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, 28, 41–66.
- Ravignani, A., Fitch, W. T., Hanke, F. D., Heinrich, T., Hurgitsch, B., Kotz, S. A., Scharff, C., Stoeger, A. S., & de Boer, B. (2016). What Pinnipeds Have to Say about Human Speech, Music, and the Evolution of Rhythm. *Frontiers in Neuroscience*, 10. <https://doi.org/10.3389/fnins.2016.00274>
- Reed, M. & Michaud, C. (2015). Intonation in Research and Practice: The Importance of Metacognition In M. Reed & J. Levis *The handbook of English pronunciation* (pp. 397-412). Wiley-Blackwell.
- Richards, J. C., & Rodgers, T. S. (2001). *Approaches and Methods in Language Teaching* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511667305>
- Ringbom, H. and Jarvis, S. (2009). The Importance of Cross-Linguistic Similarity in Foreign Language Learning. In M.H. Long and C.J. Doughty Eds.) *The Handbook of Language Teaching* (pp. 106-118). Wiley and Blackwell. <https://doi.org/10.1002/9781444315783.ch7>
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188–194. [https://doi.org/10.1016/s0166-2236\(98\)01260-0](https://doi.org/10.1016/s0166-2236(98)01260-0)

- Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988) Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research*, 71, 491–507. <https://doi.org/10.1007/BF00248742>
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27(1), 169–192. <https://doi.org/10.1146/annurev.neuro.27.070203.144230>
- Roach, P. (1982). On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73–79). Arnold.
- Roach, P. (2009). *English Phonetics and Phonology A Practical Course (4th edition)*. Cambridge University Press.
- Roach, P. (2004). British English: Received Pronunciation. *Journal of the International Phonetic Association*, 34(2), 239–245. <https://doi.org/10.1017/S0025100304001768>
- Roberts, B. R. T., MacLeod, C. M., & Fernandes, M. A. (2022). The enactment effect: A systematic review and meta-analysis of behavioral, neuroimaging, and patient studies. *Psychological Bulletin*, 148(5–6), 397–434. <https://doi.org/10.1037/bul0000360>
- Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five Personality Factors and Personal Values. *Personality and Social Psychology Bulletin*, 28(6), 789–801. <https://doi.org/10.1177/0146167202289008>
- Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional Anatomy of Language and Music Perception: Temporal and Structural Factors Investigated Using Functional Magnetic Resonance Imaging. *The Journal of Neuroscience*, 31(10), 3843–3852. <https://doi.org/10.1523/JNEUROSCI.4515-10.2011>
- Rogerson-Revell, P. M. (2021). Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions. *RELC Journal*, 52(1), 189–205. <https://doi.org/10.1177/0033688220977406>
- Rossi, M. (1979). Le français, langue sans accent ?», In I Fónagy & P. Léon (Eds.) *L’accent en français contemporain* (pp. 93-106). Didier.
- Rota, G., & Reiterer, S. M. (2009). Cognitive aspects of pronunciation talent. In G. Dogil & S. M. Reiterer (Eds.), *Language Talent and Brain Activity* (pp. 67–96). Mouton de Gruyter. <https://doi.org/10.1515/9783110215496.67>
- Rouaud, J., Huet, N., & Przewozny, A. (2022). Assessing Spoken English Performance and Self-Efficacy Beliefs in the Classroom: Some Considerations on the Value of an Interdisciplinary Embodied

Methodology for French Learners of English. *RANAM (Recherches Anglaises et Nord-AMéricaines)*, 55/2022, 87–113.

Roussel, S., & Tricot, A. (2015). Effet de l'élaboration d'hypothèses sur la compréhension de l'oral et sur les stratégies d'autorégulation de l'écoute en langue seconde : Une étude empirique. *Alsic*, 18(1), 1–19. <https://doi.org/10.4000/alsic.2788>

Roussi, M., & Cherkaoui Messin, K. (2011). L'insécurité linguistique entre pratiques enseignantes et compétences langagières : vers une redéfinition du rôle de l'enseignant de langues. In F. Dervin & V. Badrinathan (Eds.), *L'enseignement non natif : identités et légitimité dans l'enseignement-apprentissage des langues étrangères* (pp. 237-264). EME Editions.

Ruben, R. J. (1997). A Time Frame of Critical/Sensitive Periods of Language Development. *Acta Otolaryngol*, 117, 202–205.

Ryalls, J., Le Dorze, G., Lever, N., Ouellet, L., & Larfeuil, C. (1994). The effects of age and sex on speech intonation and duration for matched statements and questions in French. *The Journal of the Acoustical Society of America*, 95(4), 2274–2276. <https://doi.org/10.1121/1.408639>

Sabatier, J. M., & Ekimov, A. E. (2008). A Review of Human Signatures in Urban Environments Using Seismic and Acoustic Methods. *2008 IEEE Conference on Technologies for Homeland Security*, 215–220. <https://doi.org/10.1109/THS.2008.4534452>

Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, 21(5), 589–608. <https://doi.org/10.1177/1362168816643111>

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Using Listener Judgments to Investigate Linguistic Influences on L2 Comprehensibility and Accentedness: A Validation and Generalization Study. *Applied Linguistics*, amv047. <https://doi.org/10.1093/applin/amv047>

Sakai, M., Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187-224. <https://doi.org/10.1017/S0142716417000418>

Sakach, A. N. (2022). Blue Canoe. *CALICO Journal*, 39(2), 248–256. <https://doi.org/10.1558/cj.18393>

Sammler, D., Grosbras, M.-H., Anwender, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and Ventral Pathways for Prosody. *Current Biology*, 25(23), 3079–3085. <https://doi.org/10.1016/j.cub.2015.10.009>

- Saussure, F. de. (1916). *Cours de linguistique générale*. Payot & Rivages.
- Sauvage, J. (2019). Phonétique et didactique : Un mariage contre-nature. *Recherches en didactique des langues et des cultures*, 16(1). <https://doi.org/10.4000/rdlc.4276>
- Sauvage, J., & Billières, M. (2019). Enseigner la phonétique d'une langue étrangère : Bilan et perspectives. *Recherches en didactique des langues et des cultures*, 16(1). <https://doi.org/10.4000/rdlc.4234>
- Savatovsky, D. (2011). Disciplinarité et instances de légitimation d'un domaine de recherche: La didactique du français et des langues au CNU. *Pratiques*, 149–150, 25–40. <https://doi.org/10.4000/pratiques.1686>
- Scott, D. R., Isard, S. D., & de Boysson-Bardies, B. (1985). Perceptual isochrony in English and in French. *Journal of Phonetics*, 13(2), 155–162. [https://doi.org/10.1016/S0095-4470\(19\)30743-0](https://doi.org/10.1016/S0095-4470(19)30743-0)
- Shewell, C. (2009). *Voice work: Art and science in changing voices*. Wiley-Blackwell.
- Schindler, S., & Bublitzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex*, 130, 362–386. <https://doi.org/10.1016/j.cortex.2020.06.010>
- Schmidt, P. (2006). Le théâtre comme art d'apprentissage de la langue étrangère. *Spirale. Revue de recherches en éducation*, 38(1), 93–109. <https://doi.org/10.3406/spira.2006.1272>
- Schmidt P., (2010). Théâtre et langues vivantes en primaire. *Les Langues Modernes*, 2(2010), 30-38
- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. *Proceedings of CLaSIC 2010, Singapore, December 2-4*, 721–737.
- Schnitzer, M. L. (1974). Applied generative phonology: A methodology for teaching pronunciation. *IRAL (International Review of Applied Linguistics in Language Teaching)*, 12(1–4). <https://doi.org/10.1515/iral.1974.12.1-4.289>
- Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336–354. <https://doi.org/10.1016/j.jneuroling.2009.12.004>
- Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a Lingua Franca. *International Journal of Applied Linguistics*, 11(2), 133–158. <https://doi.org/10.1111/1473-4192.00011>

- Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal*, 59(4), 339–341. <https://doi.org/10.1093/elt/cci064>
- Sekhar, C., Patwardhan, M., & Singh, R. Kr. (2013). A literature review on motivation. *Global Business Perspectives*, 1(4), 471–487. <https://doi.org/10.1007/s40196-013-0028-1>
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10(1–4), 209–241. <https://doi.org/10.1515/iral.1972.10.1-4.209>
- Setter, J. (2019). *Your voice speaks volumes: It's not what you say, but how you say it*. Oxford University Press.
- Sewell, A. (2021). Functional Load and the Teaching-Learning Relationship in L2 Pronunciation. *Frontiers in Communication*, 6, 1-6. <https://doi.org/10.3389/fcomm.2021.627378>
- Shatz, M. (2007). On the Development of the Field of Language Development In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 387–405). Blackwell Publishing. <https://doi.org/10.1002/9780470757833.ch1>
- Sheppard, B. (2016). English Accent Coach (review). *TESL-EJ*, 20(1).
- Shintel, H., Anderson, N. L., & Fenn, K. M. (2014). Talk This Way: The Effect of Prosodically Conveyed Semantic Information on Memory for Novel Words. *Journal of Experimental Psychology*, 143(4), 1437–1442. <https://doi.org/10.1037/a0036605>
- Silverman, D. (2011). Schwa. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1–15). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781444335262.wbctp0026>
- Simpson, A. P. (2009). Phonetic differences between male and female speech: Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3(2), 621–640. <https://doi.org/10.1111/j.1749-818X.2009.00125.x>
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, 164, 77–105. <https://doi.org/10.1016/j.bandl.2016.10.004>
- Skulmowski, A., Rey, G.D. Embodied learning: introducing a taxonomy based on bodily engagement and task integration. *Cogn. Research* 3, 6 (2018). <https://doi.org/10.1186/s41235-018-0092-9>

Slevc, L. R., & Miyake, A. (2006). Individual Differences in Second-Language Proficiency: Does Musical Ability Matter? *Psychological Science*, 17(8), 675–681. <https://doi.org/10.1111/j.1467-9280.2006.01765.x>

Sockett, G. (2014). *The Online Informal Learning of English*. Palgrave Macmillan UK. <https://doi.org/10.1057/9781137414885>

Soulaine, S. (2014). Le défi d'une transposition didactique inédite : Création d'un gestographe. In J. Aden & A. Arleo (Eds.), *Actes de Colloque : 'Langues en mouvement : Didactique des langues et pratiques artistiques'*, Vol. 6 (pp 1-21). CRINI; E-CRINI. Retrieved 10 May, 2023 from <https://crini.univ-nantes.fr/publications-crini/e-crini/actes-de-colloque-langues-en-mouvement-didactique-des-langues-et-pratiques-artistiques>

Sperling, G. (1963). A Model for Visual Memory Tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 5(1), 19–31. <https://doi.org/10.1177/001872086300500103>

Sperry, R. W. (1961). Cerebral organization and behavior. *Science*, 133(3466), 1749–1757.

Steele, J. (1779). *Prosdia Rationalis: An Essay Towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols*. J. Nichols. (reprinted by Hildesheim: Georg Olms Verlag, 1971).

Stenton, A. (2009). *The inhibition of mother tongue interference in foreign language speech perception and production—A proposed solution for European university students*. In A. M. Vilas, A. S. Martin, J. A. M. González, & J. M. González (Eds.) *Research, Reflections and Innovations in Integrating ICT in Education* (pp. 297–303). Formatex.

Stenton, A. J., & Gimeno, A. (2011). Managing the monolingual mindset. SWANS: an authoring system for raising awareness of L2 lexical stress patterns and for inhibiting mother-tongue interference. *The Eurocall Review*, 18, 75–85.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>

Stocker, L. (2017). The Impact of Foreign Accent on Credibility: An Analysis of Cognitive Statement Ratings in a Swiss Context. *Journal of Psycholinguistic Research*, 46(3), 617–628. <https://doi.org/10.1007/s10936-016-9455-x>

- Storch, N. (2017). Sociocultural theory in the L2 classroom. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 69–83). Routledge.
- Sunderland, J. (1998). Girls being quiet: A problem for foreign language classrooms? *Language Teaching Research*, 2(1), 48–82. <https://doi.org/10.1177/136216889800200104>
- Sunderland, J. (2000). Issues of language and gender in second and foreign language education. *Language Teaching*, 33(04), 203–223. <https://doi.org/10.1017/S0261444800015688>
- Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, 36(2), 219–238. <https://doi.org/10.1016/j.wocn.2007.05.001>
- Swiggers, P., & Verleyen, S. (2002). Principes fonctionnels (dans l'explication) du changement linguistique. *La linguistique*, 38(2), 105. <https://doi.org/10.3917/ling.382.0105>
- Szakay, A. (2006). Rhythm and pitch as markers of ethnicity in New Zealand English. In P. Warren & C. Watson (Eds.), *Proceedings of the 11th Australasian International Conference on Speech Science and Technology* (pp. 421–426). Auckland, New Zealand: Australasian Speech Science and Technology Association.
- Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25(1), 1–24. <https://doi.org/10.1006/jpho.1996.0031>
- Tanner, M. W., & Landon, M. M. (2009). The effects of computer-assisted pronunciation readings on esl learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning & Technology*, 13(3), 51–65. <http://dx.doi.org/10125/44191>
- Taylor, A., & Manès-Bonnisseau, C. (2018). *Propositions pour une meilleure maîtrise des langues vivantes étrangère. Oser dire le nouveau monde*. (Report commissioned by the French Ministry of Education).
- Tellier, M. (2008). Dire avec des gestes. In F. Chnane-Davin & J.-P. Cuq (Eds.), *Du discours de l'enseignant aux pratiques de l'apprenant en classe de français langue étrangère, seconde et maternelle*. (pp. 40–50).
- Thomas, M. (2009). Review article: Ethical issues in the study of second language acquisition: resources for researchers. *Second Language Research*, 25(4), 493–511. <https://doi.org/10.1177/0267658309349676>
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, 4(1), 46–64. <https://doi.org/10.1037/1528-3542.4.1.46>



- Thomson, R. I. (2011). Computer Assisted Pronunciation Training: Targeting Second Language Vowel Perception Improves Pronunciation. *CALICO Journal*, 28(3), 744–765. <https://doi.org/10.11139/cj.28.3.744-765>
- Thomson, R. I. (2013). Accent Reduction. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 8-11). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405198431.wbeal0004>
- Thomson, R. (2014). Accent reduction and pronunciation instruction are the same thing. In L. Grant (Ed.), *Pronunciation Myth* (pp. 160–187). Michigan University Press.
- Thomson, R. I., & Foote, J. A. (2019). Pronunciation teaching: Whose ethical domain is it anyways? In J. Levis & C. Nagle (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 226–236). Iowa State University.
- Thomson, R. I., & Isaacs, T. (2022). Evaluations of Foreign Accented Speech: Subjective Bias or Speech Signal Characteristics? In V. G. Sardegna & A. Jarosz (Eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training* (pp. 27–44). Springer International Publishing. [https://doi.org/10.1007/978-3-030-98218-8\\_3](https://doi.org/10.1007/978-3-030-98218-8_3)
- Timmis, I. (2002). Native-speaker norms and International English: A classroom view. *ELT Journal*, 56(3), 240–249. <https://doi.org/10.1093/elt/56.3.240>
- Tomalin, B. (2010). India rising: The need for two way training. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 172–189). Continuum.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5), 675–691. <https://doi.org/10.1017/S0140525X05000129>
- Torrent, M. (2022). *Rapport du jury : Agrégation externe. Section : langues vivantes étrangères : anglais*. Ministère de l'éducation nationale, de la jeunesse et des sports. Retrieved 10 May, 2023 from <https://www.devenirensignant.gouv.fr/cid159832/sujets-rapports-des-jurys-agregation-2022.html>
- Tortora, G.; Derrickson, B. (2009). *Principles of anatomy and physiology* (12th ed.). Harper & Row.
- Trager, George L.; Smith, Henry Lee (1951). *An Outline of English Structure*. American Council of Learned Societies.
- Traunmüller, H., & Eriksson, A. (1995). The perceptual evaluation of  $F_0$  excursions in speech as evidenced in liveliness estimations. *The Journal of the Acoustical Society of America*, 97(3), 1905–1915. <https://doi.org/10.1121/1.412942>

- Trippe, J., & Baese-Berk, M. (2019). A prosodic profile of American Aviation English. *English for Specific Purposes*, 53, 30–46. <https://doi.org/10.1016/j.esp.2018.08.006>
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(4), 905–916. <https://doi.org/10.1017/S1366728912000168>
- Trubetzkoy (1939). *Grundzüge der Phonologie*. Travaux du Cercle Linguistique de Prague, 7. [Translator C. A. M. Baltaxe, *Principles of phonology*. Berkeley: University of California Press, 1969.]
- Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, 1(2), 179–195. <https://doi.org/10.1017/S0047404500000488>
- Tulaja, L. (2020). Exploring acceptability: L1 judgements of L2 Danish learners' errors. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference*, Northern Arizona University, September 2019 (pp. 197–206). Iowa State University.
- Tummillo, F. (2018). Le travail sur la prononciation et la prosodie par la pratique théâtrale en classe de langue italienne. In O. Racine (Ed.), *Entendre, chanter, voir et se mouvoir. Réflexion sur les supports employés dans la classe de langue* (pp. 27–34). Éditions des Archives Contemporaines.
- Turin-Barthier, M. (2022). *Rapport du jury : Concours externe du CAPES et CAFEP-CAPES*. Ministère de l'éducation nationale, de la jeunesse et des sports. Retrieved 10 May, 2023 from <https://www.devenirenseignant.gouv.fr/cid159884/sujets-rapports-des-jurys-capes-2022.html>
- Ultan, R. (1978). Some general characteristics of interrogative systems. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language, Volume 2, Syntax* (pp. 211–48). Stanford University Press.
- Vaissière. (2002). Cross-linguistic prosodic transcription: French versus English. In N. Volskaya, N. Svetozarova, & P. Skrelin (Eds.), *Problems and methods of experimental phonetics. In honour of the 70th anniversary of Pr. L.V. Bondarko* (pp. 147–164). St Petersburg State University Press.
- Vaissière, J. (2008). De la voix et du sens—Autour de l'œuvre de Iván Fónagy. In R. Patrick & J. Maár (Eds.), *Temps, Espaces, Langages : La Hongrie à la croisée des disciplines* (pp. 56–70). L'Harmattan.
- Vaissière, J.; Michaud (2006). A. Prosodic constituents in French: a data-driven approach. In I. Fónagy, Y. Kawaguchi and T. Moriguchi (Eds.) *Prosody and syntax* (pp. 47-64). John Benjamins.
- Vallerand, R.J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. *Advances in Experimental Social Psychology*, 29, 271-360.

- van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Paper in Linguistics*, 13(2), 201–277. <https://doi.org/10.1080/08351818009370498>
- van Maastricht, L. (2018). *Second language prosody: Intonation and rhythm in production and perception* [PhD thesis]. Tilburg University.
- Varela, F., Rosch, E., & Thompson, E. (1992). *The embodied mind: Cognitive science and human experience*. MIT Press.
- Vauclair, J., & Meguerditchian, A. (2007). The gestural origin of language and its lateralization: Theory and data from studies in nonhuman primates. In S. Kern, F. Gayraud, & E. Marisco (Eds.), *Emergence of Linguistic Abilities: From Gestures to Grammar* (pp. 43–59). Cambridge Scholars Publishing Ltd.
- Vieru, B., De Mareüil, P. B., & Adda-Decker, M. (2011). Characterisation and identification of non-native French accents. *Speech Communication*, 53(3), 292–310. <https://doi.org/10.1016/j.specom.2010.10.002>
- Ville, S. (2021). The verbotonal method in Foreign Language Acquisition. *Proceedings of the 3rd International Symposium on Applied Phonetics (ISAPH 2021)*, Universitat Rovira i Virgili, Tarragona, Spain.
- Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond g: Putting multiple intelligences theory to the test. *Intelligence*, 34(5), 487–502. <https://doi.org/10.1016/j.intell.2006.02.004>
- Vogel, I. (2000). The Acquisition of Prosodic Phonology: Challenges for the L2 Learner. Paper presented at “Structure, Acquisition, and Change of Grammars: Phonological and Syntactic Aspects”, Hamburg, Germany.
- Vogel, I., Athanasopoulou, A., & Pincus, N. (2016). Prominence, Contrast, and the Functional Load Hypothesis: An Acoustic Investigation. In J. Heinz, R. Goedemans, & H. van der Hulst (Eds.), *Dimensions of Phonological Stress* (1st ed., pp. 123–167). Cambridge University Press. <https://doi.org/10.1017/9781316212745.006>
- Voise, A-M. (2010). Enseigner la phonologie de l’anglais aux futurs professeurs du primaire. *Recherche et pratiques pédagogiques en langues* 29(2) (pp.11-24). <https://doi.org/10.4000/apliut.673>
- Voise, A-M. (2018). Corps – accords : une approche holistique et transdisciplinaire des langues à l’école maternelle. *Recherche et pratiques pédagogiques en langues* 37(2). <https://doi.org/10.4000/apliut.6345>

Volín, J. (2019). The size of prosodic phrases in native and foreign-accented read-out monologues. *AUC Philologica*, 2019(2), 145–158. <https://doi.org/10.14712/24646830.2019.23>

Vorger, C. (2012). Vous récitez ? Eh bien, slamez maintenant ! *Recherches en didactique des langues et des cultures*, 2012(9). <https://doi.org/10.4000/rdlc.2489>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What Influences Learning? A Content Analysis of Review Literature. *The Journal of Educational Research*, 84(1), 30–43. <https://doi.org/10.1080/00220671.1990.10885988>

Wang, Y.-T., Green, J. R., Nip, I. S. B., Kent, R. D., & Kent, J. F. (2010). Breath Group Analysis for Reading and Spontaneous Speech in Healthy Adults. *Folia Phoniatica et Logopaedica*, 62(6), 297–302. <https://doi.org/10.1159/000316976>

Wang, X. (2022). Segmental versus Suprasegmental: Which One is More Important to Teach? *RELC Journal*, 53(1), 194–202. <https://doi.org/10.1177/0033688220925926>

Walker, J. (1791) *A critical pronouncing dictionary and expositor of the English language: ... To which are prefixed, principles of English pronunciation ... Likewise rules to be observed by the natives of Scotland, Ireland, and London, for avoiding their respective peculiarities, and directions to foreigners for acquiring a knowledge of the use of this dictionary; the whole interspersed with observations, philological, critical, and grammatical*. Robinson.

Walker, R., Low, E.-L., & Setter, J. (2021). *English pronunciation for a global world* [OUP report]. Oxford University Press.

Walley, A. C. (2007). Speech learning, lexical reorganization, and the development of word recognition by native and non-native English speakers. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 315-330). Amsterdam: John Benjamins. <https://doi.org/10.1075/lllt.17.27wal>

Wardhaugh, R. (1970). The Contrastive Analysis Hypothesis. *TESOL Quarterly*, 4(2), 123. <https://doi.org/10.2307/3586182>

Warren, P. (2016). *Uptalk: The Phenomenon of Rising Intonation*. Cambridge University Press, Cambridge UK.

- Waterhouse, Lynn (2006). Multiple Intelligences, the Mozart Effect, and Emotional Intelligence: A critical review. *Educational Psychologist*, 41 (4), 207–225. [https://doi:10.1207/s15326985ep4104\\_1](https://doi:10.1207/s15326985ep4104_1)
- Weber, C. (2019). Interrogations épistémologiques autour de l’oralité : Quel paradigme pour la didactique de la prononciation de demain ? *Recherches en didactique des langues et des cultures*, 16(1). <https://doi.org/10.4000/rdlc.4252>
- Weiner, B. (1976). An Attributional Approach for Educational Psychology. *Review of Research in Education*, 4, 179. <https://doi.org/10.2307/1167116>
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548–573. <https://doi.org/10.1037/0033-295X.92.4.548>
- Weiner, B. (2010). Attribution Theory. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (pp. 558-563). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470479216.corpsy0098>
- Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge University Press.
- Wells, J. (2012). HappY again. *John Wells’s Phonetic Blog*. Retrieved 10 May, 2023 from <https://phonetic-blog.blogspot.com/2012/06/happy-again.html>
- Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10(2), 193–216. [https://doi.org/10.1016/S0095-4470\(19\)30957-X](https://doi.org/10.1016/S0095-4470(19)30957-X)
- Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press.
- Wharton, S., & Wolstenholme, R. (2019). Accents et insécurité linguistique en cours d’anglais : Le cas d’étudiants de LEA1. *Lidil*, 59. <https://doi.org/10.4000/lidil.6311>
- Wheatley, T., & Wegner, D. M. (2001). Automaticity of Action, Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 991–993). Elsevier. <https://doi.org/10.1016/B0-08-043076-7/01747-2>
- Wheeler, P., & Saito, K. (2022). Second Language Speech Intelligibility Revisited: Differential Roles of Phonological Accuracy, Visual Speech, and Iconic Gesture. *The Modern Language Journal*, 106(2), 429–448. <https://doi.org/10.1111/modl.12779>
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4), 501–522. <https://doi.org/10.1016/j.wocn.2007.02.003>

- Wilner, L. K. et Feinstein-Whittaker, M. (2007). *Medically Speaking RULES*. Successfully Speaking.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. <https://doi.org/10.3758/BF03196322>
- Wilson A. (in press). Language ideologies and English teaching in French higher education. *E-Rea*.
- Wilson, A. D., & Golonka, S. (2013). Embodied Cognition is Not What you Think it is. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00058>
- Wingfield, A. Wayland, S.C., Stine, E. A. L. (1992). Adult Age Differences in the Use of Prosody for Syntactic Parsing and Recall of Spoken Sentences, *Journal of Gerontology*, 47(5), 350–356, <https://doi.org/10.1093/geronj/47.5.P350>
- Wong, F. C. K., Chandrasekaran, B., Garibaldi, K., & Wong, P. C. M. (2011). White Matter Anisotropy in the Ventral Language Pathway Predicts Sound-to-Word Learning Success. *Journal of Neuroscience*, 31(24), 8780–8785. <https://doi.org/10.1523/JNEUROSCI.0999-11.2011>
- Wood, D. (2016). Willingness to communicate and second language speech fluency: An idiodynamic investigation. *System*, 60, 11–28. <https://doi.org/10.1016/j.system.2016.05.003>
- Wörtwein, T., Chollet, M., Schauerte, B., Morency, L.-P., Stiefelhagen, R., & Scherer, S. (2015). Multimodal Public Speaking Performance Assessment. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 43–50. <https://doi.org/10.1145/2818346.2820762>
- Wrembel, M. (2007). Metacompetence-based approach to the teaching of L2 prosody: practical implications. In U. Gut & J. Trouvain (Eds.). *Non-native prosody: Phonetic description and teaching practice* (pp.189-210). Mouton de Gruyter. <https://doi.org/10.1515/9783110198751.2.189>
- Yenkimaleki, M. (2017). *Effect of prosody awareness training on the quality of consecutive interpreting between English and Farsi*. [PhD thesis], Utrecht University.
- Yorkston, K.M., Strand, E.A. & Kennedy, M.R.T. (1996). *Comprehensibility of dysarthric speech*. *American Journal of Speech Language Pathology*, 5, 55–66. <https://doi.org/10.1044/1058-0360.0501.55>
- Young, R., Messum, P. (2012). *Comprendre l'apprentissage pour mieux enseigner. Introduction à la pensée de Caleb Gattegno*. Une éducation pour demain.
- Zetterholm, E., & Abelin, Å. (2019). Ungrammatical prosody does not hinder positive evaluations. In S. Calhoun, P. Escudero, M. Tabain, & Paul Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2164–2168). Australasian Speech Science and Technology Association Inc.

- Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Royal Society B: Biological Sciences*, 288(1955), 20210500. <https://doi.org/10.1098/rspb.2021.0500>
- Zhang, X., & Romero, J. (2021). Investigating the influence of sentence stress perception on foreign accent in utterances. 3rd International Symposium on Applied Phonetics (ISAPh 2021), 76–79. <https://doi.org/10.21437/ISAPh.2021-14>
- Zhuang, Y. (2015). *Evaluating the effectiveness of teaching intonation to learners in an intensive English program* [PhD thesis]. Northern Arizona University.
- Zielinski, B. (2006). The intelligibility cocktail: An interaction between speaker and listener ingredients. *Prospect*, 21(1), 22-45.
- Zielinski, B. W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, 36(1), 69–84. <https://doi.org/10.1016/j.system.2007.11.004>
- Zielinski, B. (2015). The segmental/suprasegmental debate. In M. Reed & J. Levis *The handbook of English pronunciation* (pp. 397-412). Wiley-Blackwell. <https://doi.org/10.1002/9781118346952.ch22>
- Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 40, 1-95.
- Zoghbor, W. S. (2016). A Model for Speech Processing in Second Language Listening Activities. *English Language Teaching*, 9(2), 13. <https://doi.org/10.5539/elt.v9n2p13>
- Zora, H., Riad, T., Schwarz, I.-C., & Heldner, M. (2016). Lexical Specification of Prosodic Information in Swedish: Evidence from Mismatch Negativity. *Frontiers in Neuroscience*, 10. <https://doi.org/10.3389/fnins.2016.533>

## Appendix: prosody-focused pronunciation instruction studies included in meta-study (6.2)

	short reference	focus	participants	LX	L1	time	pedagogy	CAPT (1/0)	protocol	assessment	Gain (1/0)
1	(Schnitzer, 1974)	lexical stress	n=3 ("fairly fluent" in English)	English	French	1 session	learning Chomsky & Halle's stress placement & laxing / tensing rules		pre-test & post-test (pilot study)	recording read words	1
2	(James, 1976)	intonation	n=30, student teachers, Toronto University	English	French	4 training sessions	visual display of intonation	1	3 groups (trad, display +/- feedback)	L1 listeners (rating 0/1/2)	1
3	(Gilbert, 1980)	intonation	n=36. University students: Japanese (n=11), Cantonese (n=12)	English	Cantonese & Japanese	2 weeks	active (& kazoos) & passive learning		4 tests (prosodic ID test in 3 languages & 3 learning tests (2 with control groups))	listening to prosody stripped of phonemes to recognise language	1
4	(de Bot & Mailfert, 1982)	intonation	n=27 (Dutch > English, n=15; French>English, N=12). Kodak employees, Paris	English	Dutch & French	intensive 1-week course	visual feedback of intonation, learners imitate speech samples	1	control & test groups, post-tests after 4 weeks	learners draw intonation curves of recorded speech samples	1
5	(de Bot, 1983)	intonation	n=59 (6 groups: 3 test groups, 3 control groups). University students	English	Dutch	45 or 90 minutes	visual feedback of intonation, learners imitate	1	different time practising: 1 x 45 min vs. 2 x 45 min & with or without feedback	recorded imitations judged by three listeners on a 5-point scale	1
6	(Champagne-Muzar et al., 1993)	intonation, rhythm, segmentation	n=34 (15 in test group, 19 in control group) adults	French	English	12 lessons	perception & production		control & test group, pre-test & post-test	recordings rated by L1 French listeners on 5-point scales for segments, intonation, rhythm, and global impression.	1
7	(Derwing et al., 1997)	general oral production inc. pronunciation	n=13, students, 2 - 20 years in Canada	English	mixed	12 weeks	general English, inc prosody, body language, voice quality, intonation, rhythm		pre- & post-instruction recordings	judged for comprehensibility & accent (9-point scale) & intelligibility (transcriptions), 57 L1 English listeners	1
8	(Derwing et al., 1998)	segmental / general speaking / prosody	n=48 students	English	mixed	2 weeks	different methods (teachers chose)		3 groups of 16 (segmental / global / no pronunciation instruction)	speech samples re- & post-instruction, 48 L1 Canadian English listeners (comprehensibility, accent, fluency)	1
9	(Missaglia, 1999)	prosody	n=40 (2 groups), av. age = 20,	German	Italian	20 hours	Contrastive Prosody Method (CPM) for test group, segmental training for control group		control & test group, pre-test & post-test	recordings judged by 5 L1 German listeners, awarded 18-30 marks (italian system)	1
10	(Pennington & Ellis, 2000)	prosodic cues to lexical recall	n=30, 20-35 years old, proficient	English	Cantonese	NA (20-minute tests)	language labs		2 experiments, ss listen to recordings and use prodoci cues / lexical cues to recall	Participants answers were right or wrong (recalled correctly or not)	0



	short reference	focus	participants	LX	L1	time	pedagogy	CAPT (1/0)	protocol	assessment	Gain (1/0)
11	(Hardison, 2004)	intonation	n=16, female undergraduates, L1 General American	French	English	(?) self-regulated	students listen & repeat with real-time visualisation of F0	1	pre-test & post-test	recordings rated on a 7-point scale by 3 L1 French raters	1
12	(Akita, 2005)	prosody (and assimilation, deletion, etc.)	n=64, university students (control: 17, segmental: 23, prosody: 24)	English	Japanese	4 months (12 x 90 minutes)	4-months (90x12 sessions) segmental-oriented approach vs. Prosody-oriented approach		pre-test & post-test	read sentences (rated by 2 L1 Japanese raters) and dictation	1
13	(Couper, 2006)	stress & rhythm, (pb = epenthesis affected rhythm)	n=21, NZ immigrants, largely of Asian origin, high-intermediate level, av. 32.5 years old, av. stay 2.5 years	English	mixed Asian	2 weeks	explicit input and practice with feedback		pre-test, post-test, delayed post-test	recordings & listening discrimination tests compared to pre-collected baseline data (50 similar learners)	1
14	(Tanner & Landon, 2009)	pausing, word stress, intonation	n=75, ESL learners, 17 - 54 years old, median stay in US = four months	English	variety	6 classes over 12 weeks	self-directed computer-assisted practice using Cued Pronunciation readings	1	pre-test & post-test (perception task, controlled production task, spontaneous production task)	perception task read-aloud recording (word stress & pauses counted), perceived comprehensibility (10 naïve listeners, + 2 experts)	1
15	(Alazard et al., 2010)	prosody (intonation and final stress, esp duration)	n=4. 18-40 years old, basic French (2 level A, 2 level B)	French	English	8 weeks, 2hrs per week	MVT		test group (MVT) & control group (communicative), pre- & post-test	recorded read text (same passage), rated 1-7 by 12 L1 French listeners & acoustic analysis (pauses, accents, etc.)	1
16	(Hsieh et al., 2013)	intonation, fluency, word pronunciation, and overall pronunciation	n=14, Non-major university students	English	Chinese (Taiwan)	8 hours	CAPT with "MyEnglishTutor" (+ L1 readings of texts), "shadowing technique from interpretation practice"	1	control & test group	pre-test & post test recordings analysed for intonation, fluency, word pronunciation, and overall pronunciation	1
17	(Hardison, 2013)	intonation	n=28, graduate students at an American university (TOAEFL 550+)	English	Chinese (Taiwan)	10 x 45 minutes, 2 weeks	CAPT: web-based annotation tool + video + visual displays of F0 + practice with Real-Time Pitch (RTP) in + feedback from L1 teacher	1	control & test group	speech samples from presentations pre- & post-instruction, rated by 3 L1 American English speakers, questionnaires	1
18	(Kurt et al., 2014)	intonation	n=22, low intermediate	English	Mandarin, Japanese, Spanish, Arabic	4 weeks, intensive language school	explicit prosody instruction (input, collaborative learning, focus on form, etc.)		control group & test group, pre-test & post-test (perception & production)	students listened to and read test items and answered questions	1

short reference	focus	participants	LX	L1	time	pedagogy	CAPT (1/0)	protocol	assessment	Gain (1/0)
(Carpenter, 2015)	lexical stress	n=70 (35 French & 35 English)	French / English	French / English	1 training session	perceptual fading (duration cue exaggerated, then decreased)		control group & test group for each language set	perception ABX tests (replication of Dupoux et al., 1997)	1
(Lima, 2015)	word stress, rhythm, intonation	n=12 speakers (ITAs), 178 raters	English	several L1s	4-weeks	self-directed online learning with The Supra Tutor	1	pre-test & post-test (recorded lectures)	7-minute lecture pre- & post-instruction, rated for comprehensibility by 178 L1 raters	1
(Zhuang, 2015)	intonation	n=32 university students (16=control, 16=test group)	English	Brazilian Portugese	4-week course	listening & PRAAT photo-reading exercises & no production		test group & control group, pre-test, post-test	human & acoustic analysis of intonation patterns	1
(Capliez, 2016)	stress and intonation	n=24 students, (12 = control group, 12=prosodic group, 12 z segmental group). A2 - B2, 18-21 years old.	English	French	10 x 40 minutes	choral repetotion, teacher-led		3 groups: control, prosodic training and segmental training	pilot (N=10, corntol & test groups) , then production tests & perception tests	0
(Jung et al., 2017)	lexical stress	n=57 Korean high school ss	English	Korean	4-week course	eliciting target-stress patterns during collaborative priming tasks		control, priming with task repetition, and priming with procedural repetition.	pre-test & 2 post-tests (sentence-read-aloud tasks)	1
(Saito & Saito, 2017)	word stress, rhythm & intonation	n=20 inexperienced learners (10 in each group)	English	Japanese	single-semester (12 weeks)	suprasegmental-based instructional treatment		pre-test & post-test, test group & control group	4 L1 English raters (experienced)	1
(Yenkimaleki & Van Heuven, 2019)	stress & intonation	n=36: 3 x 12 BA student interpreters (control group and 2 test groups)	English	Farsi	8-sessions	explicit rule instruction or acoustic-perceptual instruction.		pre-test & post-test, test group & control group	full TOEFL test	1
(Chen & Tian, 2020)	lexical stress	n=33 (18 in the "rules" group, 15 in the "Praat" group)	English	Taiwanese	8-sessions	test group: visualisation of F0 with PRAAT (control: explicit rules)	1	pre-test & post-test & delayed post-test, test group & control group	production & perception task, self-observation	1
(Kostromitina & Kang, 2021)	stress & intonation (pitch range), pauses, segmentals	n=75 (15 per group: beginner, intermediate and advanced)	English	Arabic & Chinese	15-week ESL immersion	(& F0 visualisation & PRAAT vs. Explicit instruction)	1	speech samples from placement & exit tests	speech analysis & statistical analysis	0

	short reference	focus	participants	LX	L1	time	pedagogy	CAPT (1/0)	protocol	assessment	Gain (1/0)
28	(Niebuhr, 2021)	pitch range, tempo, duration, pause frequency	n=60, 22 - 46 years old, 4 groups of 15.	English	varied L1s	1x1-hour session	Web-Pitcher & an instructor via Zoom	1	pre- & post-instruction recordings	acoustic analysis & PRAAT (FO, syllable duration, pauses, inter-pausal units)	1
29	Baills et al., 2022)	intonation	n=75, undergraduate, translation, interpreting, applied languages	French	bilingual Catalan/Spanish	3 x 30 minutes	"visuospatial hand gesture movements mimicking the melodic and rhythmic patterns of target sentences"		pre- and posttest: re-reading of the original text, reading of an unknown text, and spontaneous speech.	3 L1 French expert raters, 9-point scales (comprehensibility, fluency, accentedness, segmental accuracy, and suprasegmental accuracy)	1
30	(Couper, 2022)	word stress (2-syllable words)	n=18, mainly Asian, NZ immigrants, 2 groups	English	English	4 x 50 minutes	HVPT		2 groups, AB-BA design, learners recorded responses to a perception test & explained their responses	responses transcribed and graded 1-4 for accuracy	1
31	(Lascotte & Tarone, 2022)	intonation, rhythm, phrasal stress	n=7 adult learners, summer course, high intermediate - high advanced	English	mixed	28 hours over 7 weeks	7-week course, mirroring Ted Talks	1	3 rehearsed 3-minute presentations & 1 spontaneous speech video recordings over course	recordings rated by 5 L1 English experienced raters for intelligibility (0, 1, 2, 3) and delivery (0, 1, 2)	1
								<b>12</b>			<b>30</b>

*"context is all"*

Margaret Atwood (1996: 154)